Incentive and Asset Management



Dunhong Jin

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2020

To my Parents:

Hongwei Liu & Fengwei Jin

for their unending support, encouragement and love.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Thomas Noe, for his invaluable guidance and endless insight. My genuine thanks also goes to the examiners of my doctoral thesis, Prof. John Thanassoulis and Prof. Joel Shapiro.

Throughout my DPhil studies I have been benefited from the valuable discussions with many professors. I would like to express my deep appreciation to Professors Martin Schmalz, Mungo Wilson, Thomas Hellmann, Marcin Kacperczyk, Abhiroop Mukherjee, Renée Adams, Ludovic Phalippou, Ilaria Piatti, Jiri Knesl, Alan Morrison, Dimitrios Tsomocos, Iftekhar Hasan, Yao Zeng, Zhi Da, and Debrah Noe, for their generous sharing of experience and camaraderie, both in research and in life.

Moreover, I wish to thank my beloved family and dear friends, who make my days at Oxford enlightening, enjoyable and enriching. Many thanks for supporting my idealized or even ridiculous thoughts, always. You are the constant source of inspiration and love in my life.

— Dedicated to all of you and the past years —

Every beginning

is only a sequel, after all,

and the book of events

is always open halfway through.

Abstract

This thesis studies the effects of incentives, contracts, liquidity management, and information on the asset management industry.

First, the welfare of financial professionals frequently depends both on tournament rewards for relative performance (e.g., fund inflows based on fund ranking), and high-powered bonus for absolute performance (e.g., hedge-fund performance fees). It is well known that both high-powered compensation and tournament rewards engender risk taking. Understanding how the two types of rewards interact with each other and affect managers' risk-taking strategies is of great importance and the main goal of this study. Surprisingly, we show that these two sources of risk taking, tournaments and bonus compensation, mollify the risk-taking incentives produced by the other source.

Second, in global financial markets, open-end mutual funds manage a significant amount of assets. However, a key structural vulnerability of their asset management activities is the potential mismatch in investment liquidity and daily redeemability, which, during market stress periods, will lead to runs on funds. In this study, we draw attention to a new market development, alternative pricing rules, aiming to adjust funds' net asset values to pass on funds' trading costs to transacting shareholders. Using unique data on investor transactions, we find that the new rules eliminate the first-mover advantage and help funds retain investor capital during stressed market conditions.

Finally, the third study reconciles some stylized facts of asset markets: a remarkable growth in the money invested with index funds, a drop of fund delegation fees, active managers opting to be "closet indexers," and increased investor heterogeneity. Through theoretically modeling the rational asset allocation strategies of investors, I show that wealth inequality increases demand for indexing: active funds attract the wealthiest households, who are more willing to detect closet indexing. Active fund managers have to compete more aggressively through information acquisition, which lowers their excess returns and thus delegation fees.

Contents

1 Introduction and Literature Review **1**

1.1 Motivation and Background Information 1

1.1.1 Compensation and incentive 1

1.1.2 Principal–agent problem in alternative investment 4

1.1.3 Systemic risk in mutual funds 7

1.1.4 Fund flow and investors' delegation choice 10

1.2 Outline and Contribution 13

2 The Atalanta Effect: How High-Powered Compensation Reduces Risk-Taking **16**

2.1 Introduction 16

2.2 Example 22

2.2.1 Pure bonus competition 22

2.2.2 Pure rank competition 23

2.2.3 Rank and bonus competition 24

2.3 Risk Taking in Equilibrium 26

2.3.1 Formalization of the assumptions 26

2.3.2 Configurations 29

2.3.3 Upside risk taking 34

2.3.4 How low can downside risk taking go? 38

2.3.5 Bonus compensation and overall risk? 39

2.4 Robustness of the Single-Period Baseline Model 40

2.4.1 Multiple contestants 41

2.4.2 Linear compensation 41

2.4.3 Easy bonus compensation 42

2.5 Dynamic Contests 45

2.5.1 Dynamic contests–general formulation 45

2.5.2 A dynamic predation game 48

2.6 Conclusion and Implication 51

3 Swing Pricing and Fragility in Open-End Mutual Funds 53
3.1 Introduction . . . 53
3.2 Institutional Background . . . 59
3.3 Data . . . 61
3.3.1 Sample construction and measures . . . 61
3.3.2 Descriptive statistics . . . 63
3.4 Empirical Results . . . 64
3.4.1 Dilution adjustment factor across funds and time . . . 64
3.4.2 Fund flows and alternative pricing: Cross-sectional evidence . . . 66
3.4.3 Fund flows and alternative pricing: Evidence from switching funds . 68
3.4.4 Investment stability and alternative pricing . . . 73
3.4.5 When do alternative pricing rules matter more? . . . 76
3.4.6 Do alternative pricing rules affect subsequent fund performance? . . 78
3.4.7 The costs of alternative pricing . . . 80
3.4.8 Do fund companies internalize their investors' decisions? . . . 81
3.5 Conclusion . . . 82

4 The Drivers and Inhibitors of Factor Investing 83
4.1 Introduction . . . 83
4.2 Economic Setup . . . 88
4.2.1 Assets . . . 88
4.2.2 Households, managers, and information . . . 89
4.2.3 Portfolio holding strategies . . . 92
4.2.4 Equilibrium price and price efficiency . . . 93
4.3 Asset Market Equilibrium . . . 94
4.3.1 Fees and performance . . . 94
4.3.2 Entry of skilled managers and informed households . . . 97
4.3.3 Equilibria . . . 98
4.3.4 Ceiling of factor investing . . . 103
4.4 Household Heterogeneity . . . 104
4.4.1 Who favors factor investing? . . . 106
4.4.2 Effects on capital allocation . . . 107
4.4.3 Dynamics of opportunism . . . 113
4.4.4 "Alpha" . . . 115
4.5 Drivers and Effects of Rising Factor Investing . . . 116
4.5.1 Effects on return volatility and price efficiency . . . 121
4.6 Conclusion . . . 124

5 Conclusion 126

A Notations **128**

B Appendix: Atalanta Effect: How High-Powered Compensation Reduces Risk-Taking **134**

B.1 Proofs of Results 134
B.2 Multi-Contestant Equilibria 161
B.3 Dynamic Contests 168
B.4 Formal Derivations of the Generic Properties of Contest Equilibria 177
B.5 Equilibrium Configurations with Easy Bonus Packages 186
B.6 Endogenous Bonus Compensation 191
B.7 Computation of Risk-Averse Agent Welfare 195

C Appendix: Swing Pricing and Fragility in Open-End Mutual Funds **196**

C.1 Tables 196
C.2 Online Tables 211
C.3 Figures 216

D Appendix: The Drivers and Inhibitors of Factor Investing **219**

D.1 Data Evidence 219
D.2 Asset Orthogonalization 220
D.3 Proofs 222
D.4 Heterogeneous Households 237
D.5 Rich Asset Market 246

Bibliography **252**

List of Figures

1.1 Domestic financial intermediation by type of intermediary 7
1.2 External shocks and liquidity mismatch in open-end funds 8
1.3 Ownership of US equities 12

2.1 Introducing bonus packages into rank competitions 25
2.2 Equilibrium rewards in Eq1 equilibria 32
2.3 Effect of bonus compensation on risk taking 33
2.4 Equilibrium rewards in Eq2 equilibria 34
2.5 Parametric curves for equilibrium configurations 36
2.6 Subsidy effect vs. opportunity-cost effect 38
2.7 Equilibrium reward functions under an easy bonus package: Small bonus . 43
2.8 Equilibrium reward functions under an easy bonus package: Large bonus . 44
2.9 State transition dynamics 49

4.1 Wealth inequality on capital allocation 110
4.2 Comparison between homogeneous and heterogeneous market 114
4.3 Comparative statics w.r.t. households' and managers' information cost . . . 117
4.4 Comparative statics w.r.t. investors' risk appetite 120
4.5 Comparative statics w.r.t. supply uncertainty 121
4.6 Exogenous factors harming price efficiency 123

B.1 Optimal survival functions in pure rank competitions 136
B.2 Pure rank competition: Equilibrium distributions and reward functions . . 166
B.3 Rank and bonus competition: Equilibrium distributions and reward functions 167
B.4 State transition dynamics 168
B.5 Contouring plot of N with respect to (x, y) and Ψ with respect to x/y . . . 169
B.6 Convergent sequences of iterates under different starting values, v_0 176
B.7 Equilibrium configurations and parametric curves when $\mu_S > r$ 191

C.1 Daily VIX during the sample period 216
C.2 Dilution adjustment factor 216
C.3 Fund characteristics before and after the switching event 217

C.4 End investor fund flows before and after the switching event 218

D.1 Capital and fees in US delegation industry 219
D.2 Time series evolution of wealth inequality 220
D.3 Equilibrium capital allocation w.r.t. information costs 220
D.4 Beta distribution and Gini index 241
D.5 Cognitive ability inequality on capital allocation 242
D.6 Unequal society on capital allocation 244

List of Tables

2.1 Optimal bonus compensation package 40
2.2 Equilibria at different states for dynamic, static, and pure rank competition 50

A.1 Frequently used notations and variable definitions 128

B.1 Parameters for parametric example 166
B.2 Support line slopes and expected bonus payments in different configurations 193
B.3 Equilibrium mean performance in different configurations 194
B.4 Designer welfare for different configurations 194

C.1 Descriptive statistics on fund characteristics 196
C.2 Determinants of dilution adjustment factors 197
C.3 Fund flows during market stress 198
C.4 Full swing versus partial swing versus dual priced 200
C.5 Fund flows during market stress for switchers and their matched funds . . 201
C.6 End-investor flows during market stress for switchers and their matched funds 202
C.7 Flow–performance sensitivity 203
C.8 Fund exit 205
C.9 Cross-fund differences 206
C.10 End-investor flows during market stress: Role of investor characteristics . . 207
C.11 Fund flows and future fund performance 208
C.12 Tracking error and new investors 209
C.13 Pricing rules and fund portfolio adjustments 210
C.14 Swing thresholds of partial swing funds 211
C.15 Extended robustness tests 212
C.16 Summary information on switching funds 213
C.17 End-investor flows during market stress for switchers and their matched funds: Robustness tests 214
C.18 Volatility of end-investor flows 215

Chapter 1

Introduction and Literature Review

1.1 Motivation and Background Information

1.1.1 Compensation and incentive

Rank-order competition

The application of contest theory in economics is motivated by a real-world phenomenon that employees and managers care about their performance relative to peers, and sometimes are rewarded for topping their rivals' performance. Relative-performance or tournament incentive arises when, for instance, mutual funds compete for rankings in league-tables, executives compete for promotions based on peer comparisons, social status depends on relative wealth, agents have "keeping up with the Joneses" utility based on the rivals' compensation package, etc. Rank-order tournaments were first incorporated into contract and compensation through models of labor markets. For example, Lazear and Rosen (1981) model whether prize rewards based on ranking are more efficient than piece rates.

Later, researchers considered the effect of rank incentives on fund-manager behavior. Brown et al. (1996) empirically test the hypothesis that when the managers' compensation is linked to ranking, those whose portfolios are likely to end up as "losers" will adjust their risk levels differently compared to those managing portfolios likely to be "winners." Further empirical evidence follows, such as Busse (2001), Kempf and Ruenzi (2008), etc. There are also many theoretical studies addressing rank-order competition. Roussanov (2010) and Krasny (2011) model how the desire of "Getting ahead of the Joneses" makes the portfolios biased away from the aggregate ones. Roussanov (2010) finds a lower aversion to idiosyncratic risks than to correlated risks when rank matters. Chen et al. (2018) characterize the optimal strategies of more than three mutual fund managers competing

in a multiperiod winner-take-all tournament. They show that risk-taking strategies by the leader can be very different depending on the size of the lead, the relative performance of other contestants and the amount of the remaining time. Their model is also applicable to multiperiod and heterogeneous agent settings. Moreover, Dijk et al. (2014) investigate the rank-order incentive scheme through laboratory experiments. They analyze the effect of rank incentives on portfolio skewness and managers' preferences for idiosyncratic versus correlated assets.

Some of the literature considers the contest design in a dynamic setting. These studies build dynamic games in which each player determines the stopping time of a geometric Brownian motion. The player who wins the first prize is the one with the highest stopped value. Seel and Strack (2013, 2016) deduce the closed-form equilibrium outcome in the n-player symmetric or two-player asymmetric cases. The general shape of the expected stopped process is proved to be quasi-convex in drift and variance, where the drift is negative. Feng and Hobson (2014) and Feng et al. (2016) contribute to the continuous literature on rank-order contest as well. There are many possible applications of the results in the economic background, including the duopoly competition, ranking incentive of mutual fund or private equity fund managers, and job promotion contests.

The diversity of player and prize structure, as well as the link with all-pay auction has also been examined. In essence, an all-pay auction can be regarded as a multi-player contest game. Fang and Noe (2015) build a model where contestants compete for identical prizes by choosing their random performance distributions, subject only to a capacity constraint. An innovative point of Fang and Noe (2015) is that they impose no unimodal distribution restriction. They analyze the comparative statics based on different contest structure, for example, the change in contest selectivity or size. They point out that the optimal strategy is either playing safe or "win-small/lose-big," with skewed but never risk maximizing property. Strack (2016) extends Fang and Noe (2015) by allowing the outcome distribution to rely on a random state common to all competitors. This makes the model general enough to capture the dynamic financial market models. Olszewski and Siegel (2016a) identify the effort-maximizing prize distribution using a new framework. They model the limiting behavior of large contests, as the number of contestants increases to infinity. Contestants can be heterogeneous and have private information about their ability. The equilibrium can then be described through mechanism design method, whose literature has a rich set of tools.

Performance-driven bonus compensation

Besides incentives based on relative-performance, agents are also frequently compensated by a bonus reward if they outperform in absolute terms. For instance, CEOs or high-level

executives are usually payed with a bonus for their contribution. The bonus rewards are usually stocks or options when their performance exceeds a certain threshold.[1] Absolute-performance rewards are also common to executives for meeting targets or generating profit. Bonus can be really common and generous in the real world, especially when the agents' performance is publicly observable.

In asset management industry, the "option-like" compensation structure is widely adopted by hedge funds and private equity funds. A lot of studies address this common convex compensation structure, using the ideas from agency or portfolio theory, or empirical analysis.[2] More and more active mutual funds have implemented performance-related reward structures as well. Recently some hand-collected data became available on the management contracts of mutual funds (e.g., Ma et al., 2019; Lee et al., 2019). Based on the new information, it is possible to relate the compensation structure to risk taking, performance and agency conflict in mutual funds. Mutual fund compensation is already well-known for incentivizing the rank-order tournament behavior. Incorporating performance-driven bonus structures introduces another risk-shifting incentive.

The analysis in the sequel will show that both rank-based and absolute performance rewards encourage risk taking. However, they encourage different kinds of risk taking behavior. Absolute rewards tied to challenging performance targets motivate upside risk taking. Agents bear the risk of downside losses that accompanies strategies that sometimes produce extraordinary results. The more capable the agent, the smaller the downside risk the agent must assume. Thus, the risk engendered by absolute performance rewards results from the quest of the most capable for extraordinary results. An extensive body of theoretical and empirical research suggest that, when absolute performance rewards are challenging, i.e., the reward threshold exceeds expected performance capacity, absolute performance rewards lead to "gambling on resurrection," i.e., choosing portfolios or investment projects that entail a significant probability of ruin but promise large upside returns (Rose-Ackerman, 1991; Palomino and Prat, 2003), for example, "betting the company" on a speculative R&D project.

Risk taking engendered by relative-performance incentive has a very different character. Outperforming a rival agent even by the smallest amount leads to the same ranking as topping by a large amount. Underperforming a rival agent by a large amount leads to the same ranking as just falling short of the rival's performance. Thus optimal rank based

[1] Studies of performance-related payment of top executives and CEOs include, but are not limited to, Tosi Jr and Gomez-Mejia (1989); Jensen and Murphy (1990); Haubrich (1994); Tosi et al. (2000); Cuñat and Guadalupe (2009); Frydman and Saks (2010); Cronqvist and Fahlenbrach (2013); etc.

[2] See, for instance, Starks (1987); Goetzmann et al. (2003); Hodder and Jackwerth (2007); Agarwal et al. (2009); Panageas and Westerfield (2009); Carpenter (2000); Aragon and Nanda (2012); Janeček and Sîrbu (2012); Christoffersen et al. (2013); Buraschi et al. (2014); Drechsler (2014); Guasoni and Obłój (2016); He and Kou (2018).

strategies aim to "win-small and lose-big." Risk-taking is motivated by a desire to keep up with rivals and top them by a small amount as often as possible. The less capable the agent, the higher the price the agent must bear, in the form of downside risk, to keep up. The risk generated by rank-based rewards results from less capable agents, adopting "lose-big" strategies, which given their limited capability, give them the best chance of just topping their rivals.

1.1.2 Principal–agent problem in alternative investment

A principal–agent problem involves finding the optimal contracting between, for example, an investor and a portfolio manager. According to whether the information is complete and symmetric or not, we have three cases in the contracting literature: (i) risk sharing, with symmetric information sharing which induces the first-best solution; (ii) hidden action, with unobservable or non-contractable property which induces the second-best solution; and (iii) hidden type, which under certain conditions induce the third-best solution.

In a discrete multiperiod principal–agents problem, the optimal compensation scheme is found by Holmstrom and Milgrom (1987) to be a linear function of an aggregate account. DeMarzo and Sannikov (2006) extend the discrete-time model of DeMarzo and Fishman (2004), in which the authors assume that the agent (entrepreneur) may distort the cash flow by hidden action, while the principal (investor) control the termination of the project. Then they implement it with different capital structure, e.g., credit line, long-term debt, and equity, over which the agent controls the payout policy. Sannikov (2008) derives a stochastic continuation value of an agent to describe the principal–agent problem. The continuation value has a drift and volatility determined by the agent's unobserved effort, and depends on the triggered events such as outside option, replacement and promotion. He (2009) focuses on a changing firm size, and examines both the patient and impatient agents.

Hedge funds

Hedge funds differ from the investment management firms and other types of funds along many dimensions. First, the pay of hedge fund managers is highly correlated with their performance. Second, unlike mutual funds with diversified-portfolio requirements, hedge funds may hold large percentage stakes in individual companies. Third, hedge funds can require to "lock-up" the money of investors for a period of several years. Fourth, hedge funds are permitted to trade almost all financial instruments, take short positions and trade on margin. Their use of leverage is not capped by regulators. Fifth, compared to pension funds, hedge funds generally are not subject to heightened fiduciary standards, such as

those embodied in ERISA or "prudent man" investing standards. Sixth, the majority of investors in hedge fund industry are wealthy individuals and large institutions, and money is usually raised through private offering. Finally, hedge funds suffer fewer conflicts of interest, because they are not beholden to any management concern of the companies whose shares they hold.

Hedge funds share a typical "2 and 20" compensation structure. The 20% *performance fees* are only paid on the funds' performance in excess of a "benchmark"—*high-water mark* in the industry norm. Note that in reality, the high-water mark is usually adjusted according to the hedge fund's own past performance, or the past performance of a chosen benchmark index. This convex compensation structure is necessary, in that it prevents the incapable managers from earning money by simply investing in treasury bill, known as the "fly-by-night" phenomenon as described in Axelson et al. (2009). An interesting empirical observation from Agarwal and Naik (2004) is that a hedge fund manager's overall risk-taking strategy is similar to writing an out-of-money put option on the S&P500. This is equivalent to selling an insurance to the investors. The rationale behind this phenomenon lies in that this strategy will only crash when the market is extremely bad, whereas in good years it guarantees nice payment to the managers, provided that the hedge funds have a yearly-paid incentive structure.

The convex option-like compensation, by conventional wisdom, would induce dramatic risk-taking and skewed performance distribution (e.g., Carpenter, 2000; Ross, 2004). However, in reality, hedge funds seldom take risk dramatically like this. The moderation of risk-taking has also been proved in different manners in the literature. Basak et al. (2007) show that managers exhibit an interior extremum and even decreases risk when the option is deep out-of-money. Lan et al. (2013) prove the existence of endogenously risk aversion and leverage consideration. Panageas and Westerfield (2009) and Guasoni and Obłój (2016) show that even risk-neutral managers will reduce their risk, due to their long-horizon consideration, far from Merton (1969)'s solution of unbounded risk taking. Empirically, Aragon and Nanda (2012) track performance by the extent to which fund investors' assets are below the high-water mark. They conclude that in the following situations, the risk-taking behavior of hedge funds is substantially weaker: (i) high-water-mark provision; (ii) low probability of fund liquidation; (iii) managerial stake of managers; and (iv) backfill-free period.

Another crucial component of the managers' income comes from the 2% *management fees*. Driven by it, a hedge fund manager would want to gain the maximal amount of asset under management through competing with her rivals. In fact, the importance of management-fee incentives is stronger in the current environment where topping the high-water mark is more difficult. If a rival manager only cared about beating the high-water mark, and

employed a "short-put" strategy, it would be cheap for her rival to top her in the case of market decline, by buying cheap option insurance.

Theoretically, Starks (1987) concludes that "symmetric" fee dominates the "bonus" fee from an agency theoretic view, and shows that, when the fee structure is symmetric, managers moderate their risk-taking. Lan et al. (2013) claim that management fees represent approximately 75% of the total compensation. Drechsler (2014) examines the situation when the management fees are high, which leads managers into "derisking." Empirically, Lim et al. (2015) suggest that "indirect" incentives encourage the managers in two manners: enlarging the asset base thus the management fees, and increasing the asset value of existing investors towards the high water-mark.

Private equity funds

Private equity funds invest capital in target companies called portfolio companies, through levering to sponsor a takeover and sitting on their board to manage operations. Institutional investors, like pension funds, provide capital and become limited partners (LPs). Private equity firms arise as general partners (GPs) and organize the fund as limited partnerships.

The main compensation structure of private equity industry is the "2 and 20" package, similar to other alternative investment funds, with 2% fixed *management fees* levied on the committed capital annually and 20% *carried interests* when performance exceeds a certain hurdle. What makes this industry different is the *catch-up* clause, specifying that within a small region directly above the hurdle, all the revenue should go to the GPs. For instance, an 8% hurdle plus a 100% catch-up is commonly used in the industry, as documented by Metrick and Yasuda (2010) and Sorensen et al. (2014).

The structure of private equity (including some activist hedge funds) creates two layers of principal–agent relation: (i) one between LPs and funds GPs, and (ii) one between GPs and portfolio companies. The GPs suffer little constraint while have most of the bargaining power. Compared with the above fee payments between LPs and GPs, relatively little is known about the fees charged to portfolio companies by the GPs. Recently Metrick and Yasuda (2010) and Phalippou et al. (2018) show a strong record of monetary or non-monetary transactions in this second layer. Among the transactions, the frequently mentioned components are *transaction fees* and *monitoring fees*. The former is similar to the M&A advisory fees charged at the time of IPOs and the latter is charged as a regular compensation for the time and effort of managing the firm. Ideally these fees should at least be partly shared with the LPs. However, in some detailed analysis of fee payments

in private equity (e.g., Metrick and Yasuda, 2010; Phalippou et al., 2018), the GPs enjoy some "perks" rather than directly increasing the fees specified in the contracts.

These phenomena call for investigating the compensation and incentives of private equity funds. Traditionally, at the inception of the funds, the Limited Partnership Agreements (LPAs) are signed between LPs and GPs, specifying the major fees payment including the management fees, carried interests, fractions being rebated, and some "clawback" clauses by LPs transferring some bargaining power to them. LPAs last for the entire life of the fund (10 to 15 years). However, there might exist unforeseen situations during the fund operation period, which are not able to be fully addressed *ex-ante* in the LPAs. Moreover, LPAs do not include the fees charged by the GPs to the portfolio companies. Hence, there arises the Management Services Agreements (MSAs), mostly regarded as the *ex-post* LPAs. MSAs are signed between the GPs and the delegate of the company when transaction happens. Because part of the fund managers' performance will be released when signing the MSA, they would be "penalized" for their poor performance and face difficulty raising the new capital. Therefore, how to design a socially beneficial contract to mitigate the agency cost is of great importance to the LPs in the private equity industry.

1.1.3 Systemic risk in mutual funds

In global financial markets, open-end mutual funds manage a significant and growing amount of assets, as can be seen from Figure 1.1. Open-end mutual funds are different from banks, in that they do not promise a fixed return, but rather pay according to a floating-NAV model.

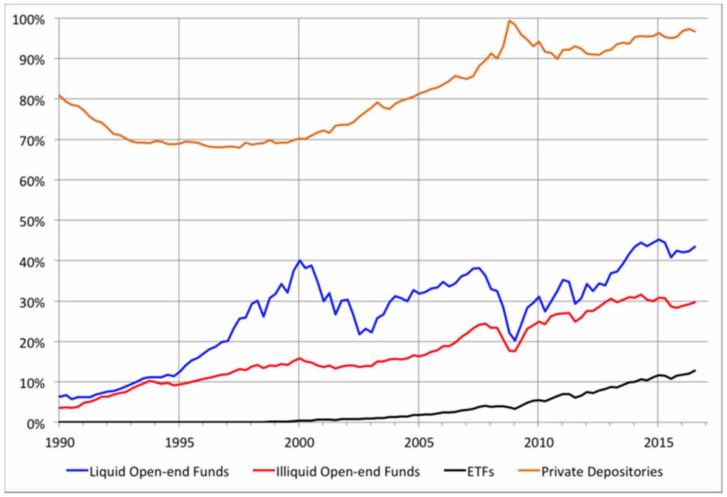


Figure 1.1: *Domestic financial intermediation by type of intermediary.* Data is from Cecchetti and Schoenholtz (2016).

A structural characteristic of open-end mutual funds is to provide liquidity transformation, because open-end funds channel investors' capital into assets that are less liquid than cash while providing daily redeemability. A consultative document "Proposed Policy Recommendations to Address Structural Vulnerabilities from Asset Management Activities" (Financial Stability Board, 2016) provides a description of the vulnerability of mutual funds to liquidity risk caused by mismatching:

"A key structural vulnerability from asset management activities is the potential mismatch in open-ended funds between liquidity of fund investments and daily redemption of fund units. [...] Investors may make significant redemptions from underperforming funds to minimize further negative returns. Funds' sale of portfolio assets required to meet these redemptions could result in greater market volatility with the potential to result in negative spillovers. [...]"

As mentioned by Bank of England in 2016 as well, the key risk at the center of the international debate is that:

"[...] Large-scale redemptions could result in sales of assets by funds that test markets' ability to absorb them without having a large impact on prices. [...]"

The main idea highlighted can be understood by a simple model: A shock leads to a reduction in asset values, investors then redeem shares in response to poor performance of their funds which forces fund managers to sell assets which then depresses asset values even further and leads to another round of redemptions, and so on. In this way, the asset management sector can lead to the creation of *systemic risk*. The resulting reduction in asset prices may ultimately have a negative impact on the real economy through higher costs of capital. The model is depicted in Figure 1.2.

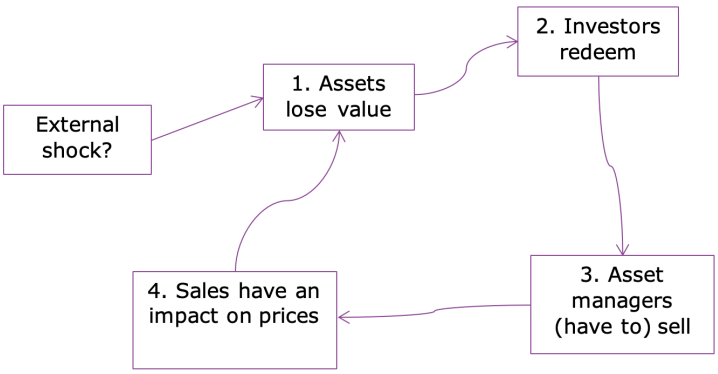


Figure 1.2: *External shocks and liquidity mismatch in open-end funds.*

Vulnerabilities arising from liquidity mismatch can be exacerbated due to the pricing mechanism used by open-end funds (e.g., Goldstein et al., 2017; Feroli et al., 2014). Under the traditional pricing mechanism, the price that a transacting shareholder receives typically

does not take into account the transaction costs that may arise from his/her transactions. Portfolio adjustments associated with meeting redemptions may occur over multiple days following the day redemption request submitted. Therefore, the costs of providing liquidity to redeeming investors could be borne by the remaining investors in the fund, thus potentially diluting the interests of non-redeeming shareholders. This mechanism can lead to "first mover advantage," thus creating incentives to run on funds, especially in times of market stress, that is, in times of declining market liquidity.

Existing literature confirms the inefficiency of the pricing mechanism, leading to runs on funds. Liu and Mello (2011) (for the case of hedge funds) and Chen et al. (2010) (for the case of open-ended mutual funds) argue that coordination risk limits funds' ability to take advantage of arbitrage opportunities. Following outflows funds need to conduct costly trades and the costs of these trades are not perfectly reflected in the NAV obtained by redeeming investors. The result is that the expectation that other investors will withdraw increases the incentive for each individual investor to withdraw as well. Evidence provided in Edelen (1999), Johnson (2004) and Coval and Stafford (2007) suggests that trades are costly, in particular for illiquid assets. Chen et al. (2010) and Feroli et al. (2014) find that outflows forecast fund performance, consistent with redemptions having a negative impact on the remaining fund investors.

Therefore, the increasing amount of assets under managed by open-end funds can be alarming because open-end fund structure can pose an important risk to financial stability, especially when these mutual funds invest in illiquid assets. Mutual fund managers, operating in many jurisdictions, are able to utilize certain tools for managing liquidity risk of open-ended funds.

One important example is that, in recent years, markets have observed an important innovation in the pricing mechanisms of open-end funds. This new pricing mechanism, which is typically called *swing pricing* in many jurisdictions, is a process of adjusting the funds' net asset value to pass on more of the costs stemming from shareholder transactions to shareholders associated with that activity. The primary goal of the new pricing mechanism is to protect the interests of existing shareholders from such costs. By removing the negative externalities arising from transacting shareholders' flows into and out of fund, consequently the new pricing mechanism can help reduce run risks.

Therefore, notwithstanding its possible merits as a liquidity management (anti-dilution) tool for individual funds, wide-acceptance of swing pricing amongst regulators has been, to some extent, the result of its perceived ability to help mitigate systemic risk. There exists theoretical work on swing pricing from Bank for International Settlements, Lewrick and Schanz (2017b), and Capponi et al. (2020), which discusses an optimal settlement price of the mechanism and positive effect of swing pricing on preventing fund failure.

However, there is barely any empirical work testing the effectiveness of the new pricing rule. Typically, a main knowledge gap identified so far is around the behavior of funds' end investors, which is at the heart of all the structural vulnerabilities. To fill this gap, data and empirical analysis on individual investor flows and fund holdings are required, in order to improve our understanding of end-investor behavior and of fund managers' reaction to outflows.

1.1.4 Fund flow and investors' delegation choice

Fund flow and performance

In asset management industry, risk taking and incentive structure are also linked to a fund's performance and fund flow. Chevalier and Ellison (1997) claim that the relationship between fund performance and the inflow of investments represents an implicit incentive contract. They find a non-linear pattern of returns in each quarter. Then they assess whether funds indeed react to the identified incentives, i.e., how mutual fund portfolios are in fact altered towards the end of the year. From the methodology point of view, Chevalier and Ellison (1997) apply a semi-parametric model to examine the flow–performance shape and analyze the agency conflicts between mutual fund companies, who want to increase expected fund inflow, and their investors, who aim at maximizing risk-adjusted returns. In a nutshell, the well established finding in the equity mutual fund research is that the flow–performance relationship tends to be convex (see, e.g., Chevalier and Ellison, 1999), i.e., equity funds' inflows to good performance appear to be more sensitive than their outflows to bad performance. However, Goldstein et al. (2017) find that for corporate bond mutual funds, the flow–performance relationship is concave. That is to say, the fund managers are heavily punished when they underperform.

Many of the subsequent studies focus on the links between mutual funds' performance, incentive structure, and fund inflow. Berk and Green (2004) provide an explanation for investors rationally choosing more skilled active managers, despite the lack of persistent performance. A strong relationship among past performance, fund inflow and diseconomies of scale does exist, which rules out any predictive power for future performance. Huang et al. (2007) study the effect of participation costs on the asymmetric flow–performance relationship. They assume that investors have a prior belief of managerial ability and face participation costs during investment. Compared to the high-cost mutual funds, the lower-cost peers are more sensitive to medium performance and less sensitive to superior performance.

In contrast to the mutual fund industry, where the persistence of performance is hard to detect, the persistence in the performance of private equity or venture capital is more significant and easier to measure. Kaplan and Schoar (2005) study from three aspects of private equity funds: the return, the persistence in fund performance, and the relation of fund performance to capital flows, fund size, and overall GP survival. The performance–flow curve shows a concave pattern. The authors offer one possible explanation as the underlying heterogeneity in the skills of GPs, and the concavity in the production function. Furthermore, Harris et al. (2012) investigate the phenomenon that too many private equity funds claim to be in top quartile, when using the performance metrics such as IRR and money multiple. The VCs and buyouts in fact make announcement of such performance status by choosing their vintage year or data benchmark judiciously. Hence, being a top quartile fund in no ways guarantees that a fund outperforms the public markets.

Last but not least, the flow–performance pattern of hedge funds is controversial despite extensive investigation: Agarwal and Naik (2004) find a convex pattern whereas Goetzmann et al. (2003) suggest a concave one. Ding et al. (2009) analyze this problem separately for the cases where hedge fund is alive or facing liquidation. Baquero and Verbeek (2013) basically show that the relationship varies over time, based on the different market conditions. Overall, the flow–performance pattern is flatter at one-quarter horizon than longer horizon, and the curve is not uniformly concave or convex. Some possible explanations could be the capacity constraints, or restriction of new fund inflows. These results help understand the management- and performance-fee incentive structure of hedge funds.

Investors' delegation choice

Over the recent decades, the asset management industry has witnessed a remarkable growth in passive investing (See Figure 1.3 for the delegation industry segmentation). The challenge to the old-fashioned active equity funds starts with Vanguard founder Jack Bogle's introduction of the very first index fund in 1976, followed by another threat in the 1990s—the emergence of exchange-traded funds (ETFs). Index-funds in the U.S. hit $4.27 trillion in assets in late 2019, surpassing the money managed by their stock-picking rivals ($4.25 trillion as of that date) for the first time ever.

Traditional active fund managers profit through rigorous stock selection, i.e., discovering the difference between asset market prices and fundamental values. Some active funds, though not all of them, possess the ability to generate excess returns. This capability, usually referred to as " skills," can include stock-picking, response to new events, industry specialization, etc.[3] On the contrary, the rising new vehicle of money management, passive

[3]c.f., Daniel et al. (1997); Kacperczyk et al. (2005); Kosowski et al. (2006); Mamaysky et al. (2008); Cuthbertson et al. (2008); Kacperczyk et al. (2014); Choi et al. (2016).

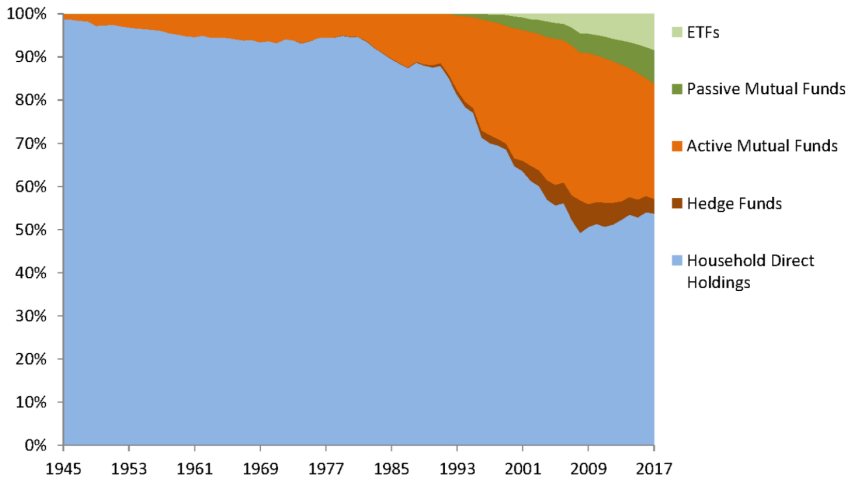


Figure 1.3: *Ownership of US equities.* Data from the Federal Reserve's Flow of Funds Report.

delegation, allows investors to buy beta exposure and hold a broadly diversified portfolio with minimal delegation fees. They have, however, scant interest or inside information towards the security-specific attributes.

Scrutiny has also been attracted towards the dark side of the growth of indexing. The traditional active funds is functioned in a way to discover price, eliminate mispricing, and monitor through their stock-picking strategies. Taken the market overall quality as given, index funds partly free-ride on the results by simply mimicking the market, thereby could possibly affect price efficiency and distort asset prices. Hence, the rise of passive investing has attracted considerable attention from practitioners, regulators and academia, on the possible impact of rising indexing on financial stability and asset prices.

One concern is the asset price informativeness. The empirical findings are a bit mixed. Israeli et al. (2017) investigate the ETF activity data and find that, over the long run ETF activity deteriorate price efficiency. However, other evidence supports an opposite view. For instance, Glosten et al. (2016) find that ETF incorporates the news of current quarter into returns in a timely manner, improving especially the price efficiency for stocks with weak information environments, high short sale constraints, or low liquidity. They also argue that ETF trading does not predict fundamentals, but only incorporates the macroeconomic news, mainly through improving the link between fundamental and asset prices.

The amplification of passive delegation will also affect volatility in general. Existing empirical papers find that the rise in passive investing increases volatility, auto-correlation, and return co-movement of assets in the index (c.f., Glosten et al., 2016; Israeli et al., 2017; Ben-David et al., 2018; Coles et al., 2018a). Among them Glosten et al. (2016) point out that the increasing co-movement can be partly driven by fundamental factors, whereas the main stream of the literature shows a non-fundamental factor-driven results

(e.g., Harris and Gurel, 1986; Vijh, 1994). Specifically, the debate of these impacts on financial stability attracts the attention from regulation and law authorities.[4]

Moreover, the rise in passive investing not only influences fund inflows into different investment vehicles, but also reciprocally determines the structure and fees inside delegation industry. Vayanos and Woolley (2016) conclude that adopting the index-based investment strategy circumvents agency problems when active managers have heterogeneous abilities. International evidence has been found by Cremers et al. (2016), that indexing adds competition pressure, forcing active funds to compete by lowering their fees.

Last but not least, there are some other bright or dark sides of passive investing. Theoretically, Malamud (2016) builds a dynamic equilibrium model and show that temporary demand shocks of ETFs have persistent impacts on future prices. Empirically, Ramaswamy (2011) links the rise of ETFs to a build-up of systematic risk. Hamm (2014) documents that ETF ownership is positively correlated with asset illiquidity.

1.2 Outline and Contribution

We start by theoretically studying the impact of combining relative- and absolute-performance rewards in Chapter 2. In the asset management industry, executives, CEOs and fund managers are widely compensated with these mixed packages. As discussed in Section 1.1, taking the incentives isolated, relative-performance or rank-order incentive suggests "win-small and lose-big" strategies and generates randomization risk, because agents always want to outperform their rivals by a tiny margin; whereas bonus incentive induces a gambling-like behavior thus leads to dramatic risk taking, as long as chasing the bonus is challenging and costly.

However, conditional on the existing research results, it is unclear how do these two incentives interact with each other to determine the risk-taking strategy of the agents. In this part of study, we aim to characterize the effect of given reward structures on behavior. We emphasize that risk-taking effects of combining rewards are not "additive": each type of reward mollifies the risk-taking incentives generated by the other. In other words, the combined package is to some extent socially beneficial. This study fills the gap of understanding the widely-adopted compensation package and its reasonability and impact.

In Chapter 3, we study the liquidity mismatch problem in open-end mutual funds. Investor "runs" on open-end funds during times of market stress can cause fire sales, even

[4]One can refer to the predictions and discussions in Sushko and Turner (2018), Anadu et al. (2018), and James et al. (2019).

destabilizing the market. This financial stability issue is of great importance, yet not well understood because people lack the data.

For this purpose, we utilize a unique and confidential dataset from Financial Conduct Authority (FCA), to study the corporate bond mutual funds—those with significant liquidity mismatch problem, and thus high run risks. As can be seen, in the post-crisis period, the growth in fixed income funds has been in multiples of the growth in equity and money market funds. Therefore, such a dataset with rich fund- and investor-level information guarantees us with comprehensive analyses.

Chen et al. (2010) argue that the culprit of run behavior is the pricing rule. In open-end funds, when investors redeem their shares at the daily-close NAV, portfolio adjustments associated with these transactions may occur over multiple days thereafter. The trading costs are then borne by non-transacting investors. This creates first mover advantage and leads to runs on open-end funds. Recently, market observes an innovation of alternative pricing rules. The goal is to protect the non-transacting investors, through adjust funds' NAV to pass on funds' trading costs to transacting shareholders.

To the best of our knowledge, this part of study is the first to draw attention to this market development and systematically examine its impacts. We aim to understand whether the new pricing mechanism alleviates incentives to runs and therefore help mitigate systemic risks. We do find that the new pricing rule eliminates the first-mover advantage and helps funds retain their investor capital during stress periods. This result is robust by including a wide range of fixed effects and the use of data on end-investors' investment decisions.

In Chapter 4, I theoretically study what determines the investors' capital allocation and thus the market shares of different types of funds in asset management industry. This part of analysis sheds light on the possible drivers of the rise in indexing.

As a consensus, a fundamental insight into the asset management industry is that we have to understand the true nature of the demand side. Most of the existing theory has a very simplistic notion of identical investors that does not reflect reality. To understand the structure of the demand, we have to take into consideration the heterogeneity or inequality among investors. This then generates a market where investors have different demand schedules, represented by, for instance, different risk preferences, stock market participation rates or delegation choices.

Currently, we observe a large degree of inequality among households investing with funds, with respect to their wealth endowment, education level, etc. On the flip side, there are two big stylized facts in the asset management industry: (i) The money invested in index funds has increased remarkably, and a large amount of this increase in indexing is accounted for by smart-beta or factor funds. (ii) The delegation fees of both index

and active funds keep decreasing. In this part of study, I reconcile these facts through a theoretical framework, taking household heterogeneity exogenous, and both households' capital allocation and fund fees endogenous.

As far as I know, this is the first study that shows that household heterogeneity will lead them to different products from asset managers, thereby affecting the endogenous capital allocation shift between different investment vehicles in general. A unique finding is that household heterogeneity serves as one possible explanation of the increase in factor investing and the decrease in both factor and active fund fees. The model further explores other possible drivers of rising factor investing, and use them to explain the stylized facts of asset market equilibria including fees, information acquisition, manager opportunism, price efficiency, and return volatility.

Chapter 2

The Atalanta Effect: How High-Powered Compensation Reduces Risk-Taking

Abstract: Agents frequently compete for both relative-performance rewards—mutual fund inflows generated by rankings, executive promotions based on peer comparisons, rank-dependent social status—and absolute-performance rewards—bonus payments for meeting targets, performance fees based on "high-water marks," wealth from portfolio returns. It is well known that bonus compensation can engender risk-taking. However, we show that introducing bonus rewards into rank-based competitions can *reduce* risk-taking through the "Atalanta effect": bonus rewards distract superior competitors from rank competition, thereby leveling the playing field and consequently reducing downside risk taking by weak competitors.

2.1 Introduction

Frequently agents' welfare depends on the ranking of their performance relative to others: mutual fund managers compete for rankings in league-tables, a key determinant of fund inflows and thus fund manager compensation (Chen et al., 2018; Chevalier and Ellison, 1997; Del Guercio and Tkac, 2008; Ma et al., 2019). Executives compete for promotions based on relative ranking (Goel and Thakor, 2008; Kini and Williams, 2012; Coles et al., 2018b). In fact an estimated 20% of large US firms use "forced ranking systems," that base promotion and termination entirely on relative performance (Bates, 2003). Ranking can also affect welfare through its effect on social status (Becker et al., 2005; Ray and Robson, 2012). Ranking concerns even drive the risk-taking decisions of financial professions when ranking is anonymous (and thus divorced from status) and has no effect on pecuniary rewards (Kirchler et al., 2018). Thus, for quite disparate reasons, a considerable body of research suggests that many financial agents are *rank-motivated*, i.e., their welfare is

affected by their ranking relative to other agents. At the same time, agent welfare is obviously also affected by non-rank absolute performance rewards, e.g., bonus compensation, stock options, portfolio values.

A literally classic example of mixed rank and non-rank rewards is provided by the mythical race between the huntress Atalanta and her suitor Melanion. When pressured by her father to marry, Atalanta agreed to marry the first suitor who defeated her in a footrace, provided that losing suitors would be executed. Atalanta was an accomplished sprinter; many suitors died. Because the outcome of these races depended purely on relative performance, these races were pure rank-based competitions.

By tossing three golden apples, Melanion introduced non-rank based incentives into this rank-based competition. Atalanta now faced a mix of rank and non-rank based rewards, escaping marriage and picking up golden apples. The pursuit of the golden apples distracted Atalanta sufficiently for Melanion to eke out a victory in the rank-based marriage competition.

In this paper we consider financial contests which, like the Atalanta/Melanion race, feature mixed rank and non-rank rewards. Specifically, we consider agents who can devote their capacities either to capturing rank-based rewards or the absolute performance rewards provided by *challenging bonus packages*, bonus compensation packages with bonus targets sufficiently high to encourage risk taking in the absence of rank rewards. Theoretical models and empirical studies have linked both bonus rewards and rank-based rewards with risk taking (e.g., Strack, 2016; Palomino and Prat, 2003; Lee et al., 2016; Ma and Tang, 2019). The question we ask is how challenging bonus packages affect risk taking by rank-motivated agents. We establish the following characterizations:

(i) *Atalanta effect:* challenging bonus packages always reduce downside risk taking by rank-motivated agents.
(ii) In some cases, challenging bonus packages also reduce upside risk taking by rank-motivated agents.

These results do not depend on the origins of rank rewards or the motivations for offering bonus compensation. Thus, this paper does not attempt to determine optimal bonus compensation polices. Rather, like Green (1984), Martynova and Perotti (2018), and Carpenter (2000), it aims to characterize the effect of given reward structures on behavior. As the literature discussed above notes, rank rewards, in practice, are generated by a variety of forces frequently outside of the control of firms. The motivations advanced for offering bonus compensation are even more varied.[1]

[1]Rationales range from managerial bargaining power (Bebchuk et al., 2001) to optimizing managerial multi-tasking (Matějka and Ray, 2017).

We do not claim that bonus compensation packages are provided by firms to reduce risk taking. We do show that bonus compensation can make contestant performance more attractive to risk averse agents (Section 2.3.5). However, bonus compensation for rank-motivated agents can be motivated by considerations orthogonal to risk reduction. In fact, in Appendix Section B.6, we endogenize bonus compensation in a framework where all agents are indifferent to the risks created by rank competition.[2] Our positive predictions on the effects of bonus compensation on risk taking are invariant to the motivations underlying the design of bonus compensation.

In our baseline, single-period analysis, two contestants with unequal ability compete by submitting *performance*. Performance is a random variable. For example, mutual fund managers' performance might consist of the returns or net asset values of their portfolios at the end of the period. Unequal ability is modeled by a *capacity* constraint, an upper bound on mean performance that is not the same for the two contestants. This specification is consistent with portfolio construction in complete markets and with the "unlimited operating flexibility" assumption used in models of corporate and venture-capital finance (Admati and Pfleiderer, 1994; Ravid and Spiegel, 1997).[3]

Contestants are also able to capture a bonus payment if their performance exceeds a threshold level. This bonus scheme, although stylized, closely tracks many standardized incentive schemes, e.g., 80/120 bonus plans which are, by far, the most common bonus compensation schemes for managers in U.S. corporations (Murphy, 1999).[4] In the baseline analysis, we assume that bonus packages are *challenging*, i.e., the contestants have insufficient capacity to capture the bonus with certainty.

In addition to bonus packages, contestants also receive rank-based rewards: they earn a rank reward whenever their performance tops the performance of their rival. We model competition for rank as a risk-taking contest. The risk-taking contest framework has been used to model rank competitions in financial market contests (Seel and Strack, 2013; Strack, 2016), social status contests (Robson, 1992; Becker et al., 2005; Ray and Robson, 2012) and political contests (Myerson, 1993). The risk taking contest framework slightly modifies the zero-noise "all-pay auction" rank-competition framework by limiting contestant performance through an expectational capacity constraint rather than through bid or

[2] In the appendix, we endogenize bonus compensation in a framework in which all agents are risk neutral. Bonus compensation is used by firms to motivate increased mean performance.

[3] It is possible to impose further restrictions on contestants' choice spaces, i.e., caps on maximum performance without substantially affecting the results of this analysis, provided that the caps are not so restrictive that they make attaining challenging bonus targets impossible.

[4] Under an 80/120 bonus scheme, the firm sets a performance target and associated maximum bonus. If the manager's performance equals or exceeds 120% of the target, the manager receives the maximum bonus; if performance falls short of 80% of the target, no payment is made to the manager; between 80% and 120% of the target, the bonus payment increases linearly in performance.

effort costs.[5] The all-pay auction framework (e.g., Olszewski and Siegel, 2016b) has been extensively analyzed and is perhaps the most widely used framework for modeling rank competition.[6] Thus, the two component parts of our framework are quite standard.

Combining these two standard frameworks to model mixed incentives produces a surprising result—the Atalanta effect: Although both bonus and rank competition, in isolation, engender downside risk taking, introducing bonus compensation into rank competitions reduces downside risk taking. Although surprising, the logic behind the Atalanta effect is fairly straightforward. In rank contests, winning small, i.e., outperforming rivals by a tiny margin, reaps the same reward as winning big; losing big produces the same reward as losing small. Downside risk taking results because weak contestants adopt "lose-big/win-small" strategies that involve downside risk taking: either mimicking strong contestants' performance or submitting the worst feasible performance, zero.

Because weak contestants' performance is low on average, it is easy for strong contestant to top with high probability. Thus, the marginal gain to strong contestants from applying capacity to attain rank dominance is smaller. Hence, when bonus packages are introduced, bonus chasing is more attractive to strong contestants. Capturing bonuses requires strong contestant to divert capacity to reaching the bonus target. Because targets are challenging, strong contestants are unable to always submit performance above the bonus threshold. Consequently, bonus chasing makes strong contestants less effective rank competitors at performance levels less than the bonus target. This reduces the amount of downside risk the weak contestant must assume to effectively compete for rank. Because only weak contestants assume downside risk, the overall level of downside risk taking falls.

The effect of bonus compensation on upside risk taking, the upper bound for the contestants' performance, is less clear cut than its effect on downside risk taking. However, we show that when bonus targets are challenging but not extremely high and either (a) strength asymmetry between the contestants is extreme or (b) bonus compensation is fairly generous, bonus compensation also reduces upside risk-taking.

[5]The alternative to the zero-noise framework is the noisy contest framework, e.g., Lazear–Rosen tournaments (Lazear and Rosen, 1981) and Tullock auctions (Buchanan et al., 1980), where performance plus an exogenous noise term determines ranking. When the noise level is small, equilibria in the noisy contest setting are fairly intractable: mixed strategies over a countable set of performance levels whose only accumulation point is zero (cf. Ewerhart, 2015, 2017; Fu and Lu, 2012). For this reason, almost all papers in the noisy contest literature assume that the noise level is sufficient to guarantee the existence of pure-strategy, no risk-taking, equilibria. For this reason, noisy contest models are not used to model rank-motivated risk taking.

[6]In fact, the risk taking and all-pay auction equilibria are related by a correspondence: the equilibrium outcomes of the risk-taking contest are the same as the equilibrium outcomes of an all-pay auction in which the bidders' valuation of the good in the all-pay auction equilibrium equals the reciprocal of the contestants' shadow price of capacity (i.e., Lagrange multipliers) in the risk-taking contest equilibrium. For more discussion, see Appendix Section B.2.

Because the Atalanta effect is founded on very generic properties of rank contests, the effect is not dependent on the specific simplifying assumptions of the baseline single-period model, i.e., the number of contestants, the exact form of the bonus contract. The implication of the Atalanta effect—reduced downside risk taking—does however depend on its hypothesis—challenging bonus packages. When bonus packages are *easy*, i.e., the bonus can be attained with probability one at least by strong contestants, bonus compensation can *increase* downside risk taking. Thus, when bonus compensation, in isolation, does not induce risk taking, it can increase risk taking when introduced into rank competitions.

Although the Atlanta effect appears to flow from quite generic properties of single-period rank competitions, it is not obvious that it is robust to the incentives produced by dynamic competition. Dynamic contests generate continuation values, values of future payoffs that depend on current actions. These continuation values might make the marginal return from rank competition higher for stronger contestants, leading stronger contestant to focus more on rank competition and thus force weaker contestants to take even more downside risk. It might seem that this continuation value effect should be especially strong when the stronger contestant can "knock out" the weaker contestant by winning one more rank competition, thereby monopolizing all future rank and bonus rewards.

We show that the continuation value effects produced in the dynamic settings can actually favor the Atalanta effect: once a contestant has a big lead, the question of which contestant will eventually emerge victorious is largely resolved. Thus, the *marginal* increase in continuation value to the leading contestant from topping a trailing rival once more can be quite small. This low marginal benefit encourages the leader to "chase the golden apples" and divert capacity to bonus chasing, thereby reducing the leader's efficiency as a rank competitor and consequently reducing risk taking by the weaker contestant.

Contributions and related literature

Our analysis makes positive predictions on the relationship between risk taking and compensation when agents compete in rank-order contests. The hypotheses tested in the extensive body of empirical financial economics research on relative performance incentives are also substantially founded on rank-order contest theory.[7] Thus, our predictions are clearly relevant to empirical research. A basic implication of our analysis for this research is that the function linking risk taking with rank and absolute performance rewards is not linear, not separable, and not always monotone. Thus, tests that aim to identify

[7]See, for example, Coles et al. (2018b); Kini and Williams (2012); Kempf and Ruenzi (2008); Chevalier and Ellison (1997); Kale et al. (2009).

the risk taking/rank competition relationship using reduced-form linear specifications to control for performance pay are misspecified.[8]

Experimental research can also benefit from a theoretical framework that encompasses the interaction between rank and absolute performance rewards. Kirchler et al. (2018), in an experiment using financial professionals, found underperforming professionals selected significantly more risky portfolios when the experimenters confidentially informed them of their relative performance rankings, even though no pecuniary rewards were associated with these rankings. Thus, Kirchler et al. (2018) suggests that agents, or at least financial professionals, will pursue rank dominance even when pecuniary rewards and social status are not tied to performance rankings.

Kirchler et al. (2018) also found that linear compensation packages had little effect on subject behavior. In our framework, linear incentive packages have no effect on risk-taking. So, at least in our setting, this result is not surprising. More importantly perhaps, our analysis can be used to calibrate non-linear compensation packages for laboratory experiments which can identify, from contestant strategies, contestants' marginal rate of substitution between monetary absolute-performance rewards and the non-pecuniary rewards of rank dominance.

Because this paper integrates incentive compensation into the analysis of asymmetric contests and tournaments, it is closely related both to research on asymmetric contests (e.g., Xiao, 2016; Siegel, 2010; Hillman and Samet, 1987; Baye et al., 1996), and research on the effects of incentive compensation on risk-taking (e.g., Drechsler, 2014; Rose-Ackerman, 1991; Van Wesep and Wang, 2014). Our results extend the contest literature by incorporating the effect of absolute performance rewards into a standard rank-competition framework, and extend the compensation literature by incorporating rank incentives.[9]

Our analysis of rank incentives is also, in a general sense, related to research based on "keeping up with the Joneses" (KUJ) utility, i.e., utility that depends on the compensation received by rival agents (e.g., Lazear, 1989; Levine, 1991; Akerlof and Yellen, 1990; DeMarzo and Kaniel, 2017; Roussanov, 2010). However, the effects of rank-based rewards and KUJ utility are quite different. For example, KUJ models cannot accommodate non-pecuniary, non status-based preferences for rank dominance (Kirchler et al., 2018). Moreover, KUJ models almost always assume a smooth relationship between an agent's utility and the compensation of rival agents. Thus, compensation just below and just above rival agents' average compensation produces approximately the same utility. For

[8]More discussion of the empirical applications of our results is provided in Section 2.6.

[9]Because, the pursuit of rank and absolute performance rewards can be thought of as multitasking, our work is also somewhat related to the multitasking literature (Holmstrom and Milgrom, 1991; Schöttner, 2007).

this reason, KUJ models, in contrast to rank-competition models, do not encourage "win small/lose big" risk taking strategies, the focus of our analysis.

2.2 Example

Before introducing the general model, we develop an example that captures the intuition behind its fundamental insight. The development of the example is heuristic. The formal analysis, which will fill in some of the technical details, commences in the next section.

Consider two rival money managers in a single-period world: a weak manager (he), W, and a strong manager (she), S. These managers trade at date 0 and buy assets that produce a liquidating dividend at date 1, this liquidating dividend is a non-negative random variable, represented by X_W and X_S, respectively. We term the realized dividend the manager's *performance*.

The risk-neutral managers receive a bonus, $b \geq 0$, if their performance weakly exceeds a bonus threshold, r. In addition, the managers may receive a rank-based reward if their performance exceeds the performance of their rival. In the event of a performance-tie, the rank reward is divided equally between the two managers. A manager's *payoff* is the sum of the manager's expected rank and bonus rewards.

Financial markets are perfect and valuation is risk neutral. This implies that the attainability of a performance distribution depends only on its expected value. Absent bonus rewards, this specification is a special case of the randomly complete financial-market contest model of Strack (2016). Thus, the only constraint on the manager's performance is that expected performance, $\mathbb{E}[X_i]$, $i = S, W$, is limited by a *capacity constraint*, an upper bound on expected performance that depends on managerial ability, termed the manager's *capacity*. We assume that the capacities of the two managers are common knowledge. The strong manager's capacity equals μ_S and the weak manager's capacity equals μ_W, where $0 < \mu_W \leq \mu_S$. We also assume that neither manager has sufficient capacity to capture the bonus with certainty, i.e., $r > \mu_S$. In our example, the weak manager's capacity, μ_W, is 1, and the strong manager's capacity, μ_S, is 3.

2.2.1 Pure bonus competition

Suppose the bonus payment, b, equals 1, and the bonus threshold, r, equals 15, and there are no rank-based rewards. Given the capacity constraint, neither managers can submit performance that always weakly exceeds the bonus threshold. Thus, both managers will choose performance distributions that maximize the probability of attaining the bonus

threshold subject to their capacity constraints. Consequently, both managers will adopt a "bang-bang" strategy of placing all probability weight on 0 or r.

In this bonus competition game, let $p_i^0(\mathrm{B})$ represent the probability of a performance of 0 for a manager of type $i = S, W$. The capacity constraint is clearly binding. Thus, optimal performance strategies satisfy

$$p_i^0(\mathrm{B}) \times 0 + \big(1 - p_i^0(\mathrm{B})\big) \times 15 = \mu_i, \quad i = S, W.$$

The assumed capacities of the managers, and the fact that $r = 15$, imply that $p_S^0(\mathrm{B}) = 4/5$ and $p_W^0(\mathrm{B}) = 14/15$. Consequently, bonus competition engenders significant downside risk taking.

2.2.2 Pure rank competition

Next, consider pure rank competition game. Specifically, assume that the bonus payment $b = 0$, and the rank reward equals 1. The equilibria of such games have been extensively analyzed in the literature (e.g., Xiao, 2016; Hillman and Riley, 1989). The equilibrium calls for the strong manager to randomize uniformly over $[0, 6]$, and the weak manager to submit a performance of 0 with probability $p_W^0(\mathrm{R}) = 2/3$ and, with probability $1 - p_W^0(\mathrm{R}) = 1/3$, also submit performance uniformly distributed over $[0, 6]$.

Verifying this equilibrium is straightforward. First, consider the weak manager. Given the uniform distribution played by the strong manager, all distributions over $[0, 6]$ featuring the same expected performance produce the same payoff. Performance in excess of 6 produces the same probability of winning the rank reward as 6, and uses up more capacity. Because the capacity constraint is binding, using more capacity to attain the same probability of winning the rank reward lowers a manager's payoff.

Similarly, for the strong manager, because the weak manager is randomizing between a point mass at 0 and a uniform distribution over $[0, 6]$, all performance distributions placing no probability mass at 0, supported by $[0, 6]$, and satisfying the capacity constraint, produce the same payoff. Any strategy placing a positive mass on 0 results in a performance tie between the weak and strong manager at 0. An arbitrarily small increase in performance breaks the tie and increases the strong manager's probability of winning from $1/2 \times 2/3$ to at least $2/3$. Since the shadow price of capacity is not infinite, a sufficiently small increase in performance increases the strong manager's payoff. Any performance level in excess of 6 produces the same probability of winning the rank contest, 1, as a performance of 6, and uses more capacity.

Thus, in this pure rank competitions, the weak manager aims to shadow the performance of the strong manager to the extent he is able and compensates for his inferior capacity by taking downside risk. The capacity difference between the two managers is reflected only in difference between their downside performance levels.[10]

2.2.3 Rank and bonus competition

Now add a bonus package to the rank-based competition. Thus, assume that, in addition to the rank reward of 1, the managers receive the same bonus compensation as they received in the pure bonus competition game: a bonus payment $b = 1$ with a bonus threshold $r = 15$.

Our first observation is that the pure rank competition equilibrium strategies are not equilibrium strategies in this rank and bonus competition. A necessary condition for the pure rank competition strategies to be best replies in the rank and bonus competition is that the strong manager cannot gain from deviating to *chasing the bonus*, i.e., choosing a strategy that captures the bonus reward with positive probability.

If the weak manager plays his pure rank competition equilibrium strategy, all possible strong-manager payoffs from randomizing between two performances—(a) a performance between 0 and 6, the boundaries of the support of her pure rank competition equilibrium distribution, and (b) a performance of $r = 15$—lie on a line in performance/reward space whose endpoints are $(0, 2/3)$ and $(15, 2)$, where the first component of these vectors is performance and the second is the associated reward. We call this line the *bonus line*. This characterization follows because, as performance approaches 0, the strong manager's probability of winning the rank contest approaches $p_W^0(\mathrm{R}) = 2/3$, and the manager does not capture the bonus. Thus, the manager's reward equals $2/3$. If the strong manager's performance equals 15, she captures the bonus $b = 1$, and, because the highest performance under the pure rank competition strategies is 6, she also captures the rank reward, 1. Thus, her reward equals $1 + b = 2$. The slope of the bonus line, which we represent by Δ_S^{Bonus}, equals

$$\Delta_S^{\text{Bonus}} = \frac{2 - 2/3}{15 - 0} = \frac{4}{45} \approx 0.089. \tag{2.2.1}$$

Similarly, all possible rewards to the strong manager from randomizing between two points in $[0, 6]$, lie on the line, termed the *rank line*, which connects $(0, 2/3)$ and $(6, 1)$. Thus, the

[10] In fact, conditioned on winning the rank competition, the weak manager's performance stochastically dominates the strong manager's performance. See Noe (2020) for verification of this assertion in an all-pay effort-bidding auction setting.

slope of the rank competition line, which we represent by Δ_S^{Rank}, is given by

$$\Delta_S^{\text{Rank}} = \frac{1 - 2/3}{6 - 0} = \frac{1}{18} \approx 0.056. \tag{2.2.2}$$

For the strong manager, the slope of the bonus line is greater than the rank line and the two lines have the same intercept. Consequently, all randomizations between a point in $[0, 6]$ and $r = 15$, produce a higher payoff to the strong manager than randomizations between points in $[0, 6]$, implying *a fortiori* that when both randomizations satisfy the strong manager's capacity constraint with equality, chasing the bonus produces a higher payoff than pure rank competition. Thus, when the bonus is introduced, the pure rank-competition equilibrium cannot be sustained. As illustrated in Figure 2.1, under the managers' pure rank competition equilibrium strategies, the rank line is "too flat" to make rank competition attractive to the strong manager.

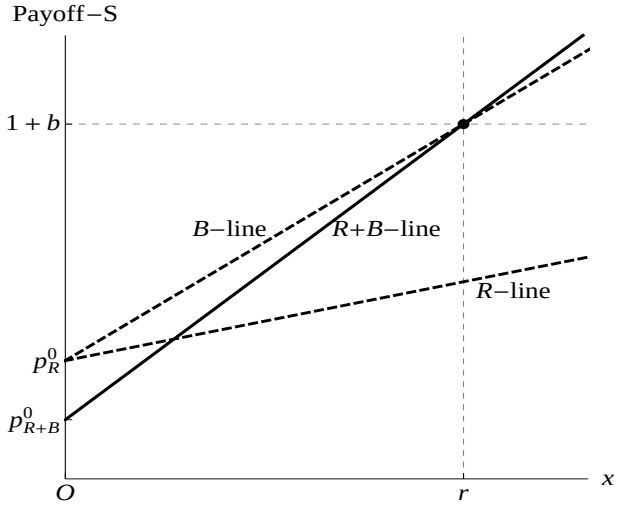


Figure 2.1: *Introducing bonus packages into rank competitions.*

However, an equilibrium in which the strong manager chases the bonus and completely ignores rank competition also cannot be sustained: if the strong manager devotes all capacity to bonus chasing and thus randomizes between 0 and $r = 15$, given the capacity constraint $\mu_S = 3$, the strong manager will win the rank and bonus competition with probability $1/5$ and otherwise receive 0, yielding a payoff of $2 \times 1/5 = 0.40$. The weak manager would, in this case, win the rank competition with probability $4/5$ and would have capacity to spare to chase the bonus, implying a much higher payoff for the weak manager, which is not possible.[11] So, in equilibrium, the strong manager must both compete for rank dominance and chase the bonus.

[11]It is always feasible for the strong manager to simply copy the performance distribution selected by the weak manager, which will result in the strong manager receiving the same payoff as the weak manager.

Let $p^0_W(\text{R+B})$ represent the probability the weak manager submits 0 performance in the rank and bonus competition equilibrium. For the strong manager to both chase the bonus and compete for rank dominance, the managers' strategies must adjust to equate the strong manager's rank and bonus lines. If, in the rank and bonus competition equilibrium, the weak manager does not chase the bonus, the rank line is determined by two points in performance/reward space—(i) the point associated with zero performance, $(0, p^0_W(\text{R+B}))$, and (ii) the point associated with performance equal to bonus threshold, $(r, 1+b) = (15, 2)$.

Given that the second point, $(r, 1+b) = (15, 2)$, is fixed, and that the slope of the rank line must increase to make rank competition attractive to the strong manager, in the rank and bonus competition equilibrium, $(0, p^0_W(\text{R+B}))$ must lie below $(0, p^0_W(\text{R}))$. Consequently, in the rank and bonus competition equilibrium, the probability of zero performance by the weak manager, $p^0_W(\text{R+B})$, must be lower than $p^0_W(\text{R})$, the probability under the pure rank competition. This adjustment is illustrated in Figure 2.1.

Using the tools developed in the next section (Lemma 2.3.3), we can compute the equilibrium strategies of the two managers: with probability $p^r_S(\text{R+B}) = 1/5 - 4/5\sqrt{31} \approx 0.06$, the strong manager targets the bonus, i.e., sets $x = r$ and, with probability $1 - p^r_S(\text{R+B})$, uniformly randomizes over $[0, u_{R+B}]$ where $u_{R+B} = \sqrt{31} - 1 \approx 4.57$; The weak manager, with probability $p^0_W(\text{R+B}) = 1/15\,(14 - \sqrt{31}) \approx 0.56$ sets $x = 0$ and, with probability $1 - p^0_W(\text{R+B})$, also uniformly randomizes over $[0, u_{R+B}]$.

Thus, as expected, $p^0_W(\text{R+B}) < p^0_W(\text{R})$, i.e., the probability of downside risk taking, is reduced by the introduction of bonus compensation. The reduction in downside risk taking results because the strong manager diverts $p^r_S(\text{R+B}) \times r \approx 0.84$ of her capacity, $\mu_S = 3$, to bonus chasing. This diversion reduces the upper bound for the strong contestant's performance below the bonus threshold from $u_R = 6$ to $u_{R+B} \approx 4.57$. Because of this reduction, the downside risk, $p^0_W(\text{R+B})$, the weak manager must assume to compete for rank dominance falls.

2.3 Risk Taking in Equilibrium

2.3.1 Formalization of the assumptions

As discussed in Section 2.2, contestants compete for both rank and bonus rewards, where performance is represented by the realization of random variables, X_S and X_W. Contestants' strategies can be characterized by the *performance distributions* they pick, denoted by F_i ($i = S$ or W). We represent the supports of the performance distribution of S and W by Supp_S and Supp_W respectively. As in the example in the previous section, we

assume the bonus threshold, r is greater than the strong contestant's capacity, μ_S, the bonus payment, b, is greater than 0, and the rank reward equals 1.

Remark 2.3.1. A very standard result in the all-pay and risk taking literature is that, in equilibrium, contestants never submit performance distributions that result in a positive probability of tied performance. If the contestants tie with positive probability, a contestant could shift probability mass from the tie point to a higher performance level arbitrarily close to the tie point. This shift would cause the contestant's payoff to jump up, while the capacity used to make the shift could be made arbitrarily small. In Appendix Section B.4, we provide a derivation of this result in the context of our model.

Because the rank-based reward is normalized to 1, and, per Remark 2.3.1, tied performance cannot occur in equilibrium, the expected rank-based rewards given performance x for contestant S and W are given by

$$\mathbb{P}(x > X_W) = \mathbb{P}(x \geq X_W) = F_W(x), \quad \mathbb{P}(x > X_S) = \mathbb{P}(x \geq X_S) = F_S(x),$$

respectively. The non-rank compensation for the contestants takes the form of a *bonus package* consisting of a bonus payment b and a bonus threshold r, where $b > 0$ and the bonus package is challenging, i.e., $r > \mu_S$. The bonus payment, b, is captured if and only if performance weakly exceeds the bonus threshold, r. Thus, the manager's bonus reward given performance x is given by $b\,\mathbb{1}_r(x)$. In this definition and in the sequel, for any real number c, we represent the indicator function for the set $[c, \infty)$ by $\mathbb{1}_c$, i.e.,

$$\mathbb{1}_c(x) = \begin{cases} 0 & x < c, \\ 1 & x \geq c. \end{cases}$$

We denote the reward from submitting a performance level of x to the strong and weak contestant by $\Pi_S(x)$ and $\Pi_W(x)$, and term Π_S and Π_W the *contest reward functions*. The reward functions are given by

$$\Pi_S(x) = F_W(x) + b\,\mathbb{1}_r(x), \quad \Pi_W(x) = F_S(x) + b\,\mathbb{1}_r(x). \tag{2.3.3}$$

Because contestants are risk neutral, their payoffs are given by the expectation of the rewards produced by the performance distributions. Thus, the payoffs to the two contestants are given by

$$\int_{\mathbb{R}^+} \Pi_i(x)\,\mathrm{d}F_i(x), \qquad i = S, W.$$

A contestant's problem can be formulated as choosing performance distribution over the non-negative real line to maximize her payoff. Each contestant faces a *capacity constraint*

which restricts her choice of performance distribution. The capacity constraint for contestant $i = S, W$, is given by

$$\int_{\mathbb{R}^+} x \, \mathrm{d}F_i(x) \leq \mu_i, \quad i = S, W.$$

We assume that the capacity, μ_i, of the strong contestant exceeds the capacity of the weak contestant, i.e., $\mu_S > \mu_W$. [12]

A pair of performance distributions, (F_S^*, F_W^*), is an equilibrium if the probability of tied performance under (F_S^*, F_W^*) is zero and each performance distribution is a best reply to the other, i.e.,

$$\int_{\mathbb{R}^+} \Pi_i(x) \, \mathrm{d}F_i^*(x) = \max \left\{ \Pi_i(x) \, \mathrm{d}F_i(x) : F_i(0-) = 0 \text{ and } \int_{\mathbb{R}^+} x \, \mathrm{d}F_i(x) \leq \mu_i \right\}, \; i = S, W, \text{ where}$$
$$\Pi_S(x) = F_W(x) + b \, \mathbb{1}_r(x), \; \Pi_W(x) = F_S(x) + b \, \mathbb{1}_r(x).$$

In Appendix Section B.4, we show that this definition identifies the same set of equilibria as would be identified if distribution pairs permitting tied performance were permitted and a tie-breaking rule was imposed that divided the rank reward equally between the contestants.

Our aim is to characterize equilibrium contestant performance distributions. Our characterization builds on a large literature on rank competitions and somewhat smaller but still significant literature on risk-taking rank contests. In order to avoid exhausting readers' patience, we will start by stating, in Remark 2.3.2, a few fairly obvious, known restrictions that these frameworks place on performance distributions. In Appendix Section B.4, we provide formal derivations of these restrictions.

Remark 2.3.2.

(a) With the possible exception of performance $x = r$, the supports of the contestants' performance distributions will coincide.
(b) $\text{Supp}_i \cup [0, r)$ and $\text{Supp}_i \cup [r, \infty)$, $i = S, W$, are connected.
(c) Contestant performance distributions are continuous except perhaps at 0 and r.
(d) 0 must be in the support of both contestants' performance distributions.

The logic behind these restrictions is that, when small changes in positive performance

[12] This assumption rules out the symmetric capacity, $\mu_S = \mu_W$. The relation between model parameters and equilibrium strategic behavior in our analysis is continuous. So, substituting μ for μ_S and μ_W recovers characterizations of equilibrium strategies in the symmetric case. However, because there is no downside risk taking in the symmetric case, all of our strict characterizations of downside risk taking with asymmetric capacity are weak and trivial in the symmetric case. We do not think it is worthwhile to expend space to provide such characterizations in our named results.

levels do not affect the probability of receiving the bonus reward, contestants' incentives are determined by the following characteristics of rank competitions: first, in rank competitions, topping rivals' performance by a small amount garners the same reward as topping by a large amount, and topping by a large amount uses more capacity. Thus, when competing for rank dominance, contestants "shadow" each other's performance. Second, an increase in performance not accompanied by an increase in the probability of topping rivals is never optimal. So over regions where the rank reward does not vary, the supports of the contestant's performance distributions are connected. Third, because discontinuities in rivals' performance strategies imply an infinite marginal gain from arbitrarily small increases in a contestant's performance, such discontinuities cannot occur in equilibrium.

Remark 2.3.3. The basic tool for identifying equilibrium distributions is a multiplier characterization of best responses. This characterization shows that the rewards produced by all performance levels, x, included in the support of a contestant's best reply are collinear. In other words, here exist *multipliers*, $\alpha_i \geq 0$ and $\beta_i > 0$, such that, if x is in the support of i's equilibrium performance distribution, then $\Pi_i(x) = \alpha_i + \beta_i\, x$. We call the map $x \hookrightarrow \alpha_i + \beta_i\, x$ contestant i's *support line*. This implies that if x' and x'' are two performance levels in the support of i's performance distribution, $\Pi_i(x'') - \Pi_i(x') = \beta_i\,(x'' - x')$, i.e., for a given contestant, the marginal gain from submitting any two performance levels in the support of the performance distribution is constant and equals to β_i. The multiplier characterization simply generalizes the marginal slope condition used to develop the example in Section 2.2. Since submitting a constant performance distribution equal to capacity is always feasible, the expected payoff to a contestant equals $\alpha_i + \beta_i\, \mu_i$. In Appendix Section B.4, we provide a formal justification for this approach.

2.3.2 Configurations

Configuration Eq0: Neither S nor W chases the bonus

Eq0 is the simplest configuration. In this configuration, the contest reduces to a purely rank-based contest. In Eq0, the bonus is not attractive to either contestant and both contestants play the pure rank-based competition strategies illustrated in Section 2.2.2, i.e., the same strategies that they would play in the absence of bonus compensation. Lemma 2.3.1 characterizes equilibrium strategies in this configuration. The proofs of this lemma and all subsequent results is deferred to Appendix Section B.1.

Lemma 2.3.1 (**Eq0: Neither contestant chases the bonus**) *When neither S nor W chases bonus, Eq0, the unique equilibrium can be characterized by the following perfor-*

mance distributions:

$$F_S^* = \text{Unif}[0, 2\,\mu_S],$$
$$F_W^* = \left(1 - \frac{\mu_W}{\mu_S}\right) \mathbb{1}_0 + \frac{\mu_W}{\mu_S}\,\text{Unif}[0, 2\,\mu_S].$$

Both contestants' capacity constraints bind, and the weak contestant's probability of downside risk taking equals $(\mu_S - \mu_W)/\mu_S$.

In this equilibrium configuration, the strong contestant randomizes uniformly over $[0, 2\,\mu_S]$. The weak contestant does not have sufficient capacity to emulate this performance distribution. The weak contestant mimics the strong contestant's performance distribution to the extent his capacity permits. This requires "paying" for high performance associated with the strong contestant by putting probability mass at 0, i.e., taking downside risk. The support line for the strong contestant is $\alpha_S + \beta_S\,x$, where $\alpha_S = (\mu_S - \mu_W)/\mu_S$ and $\beta_S = \mu_W/(2\,\mu_S^2)$ and, for the weak contestant, $\beta_W\,x$ with $\beta_W = 1/(2\,\mu_S)$. Note that, because $\mu_W/\mu_S < 1$, the marginal value of capacity for the strong contestant, β_S is less than the marginal value for the weak contestant, β_W.

Configuration Eq1: Only S chases the bonus

As b increases or r decreases, bonus chasing could become attractive for one of the contestants. Our first result is when only one contestant chases the bonus, then the bonus chasing contestant must be the strong contestant.

Lemma 2.3.2 *There do not exist equilibria in which the weak contestant chases the bonus (i.e. the probability that the weak contestant captures the bonus is positive) and the strong contestant does not chase the bonus.*

The logic for this result is simple. If the strong contestant ignores the bonus and simply pursues rank competition, the strong contestant must be applying more capacity to rank competition than the weak contestant, implying that the marginal gain from applying capacity to rank competition is lower for the strong contestant. The marginal gains from chasing the bonus are the same for both weak and strong contestants. Thus, whenever bonus chasing is attractive to the weak contestant, it is also attractive to the strong contestant.

Our next result, Lemma 2.3.3, specifies equilibrium strategies for the two contestants when only one contestant, S, chases the bonus.

Lemma 2.3.3 (**Eq1: S but not W chases bonus**) *When only S chases the bonus, the equilibrium can be characterized by the following performance distributions:*

$$F_S^* = p_S^r \, \mathbb{1}_r + (1 - p_S^r) \, \text{Unif}[0, u],$$
$$F_W^* = p_W^0 \, \mathbb{1}_0 + (1 - p_W^0) \, \text{Unif}[0, u],$$

where

$$u = \frac{2\,r\,\mu_W}{\mu_W + \sqrt{2\,b\,r\,\mu_W + \mu_W^2}} < r, \qquad p_S^r = \frac{2\,\mu_S - u}{2\,r - u} \in (0,1), \qquad p_W^0 = 1 - \frac{2}{u}\,\mu_W \in (0,1).$$

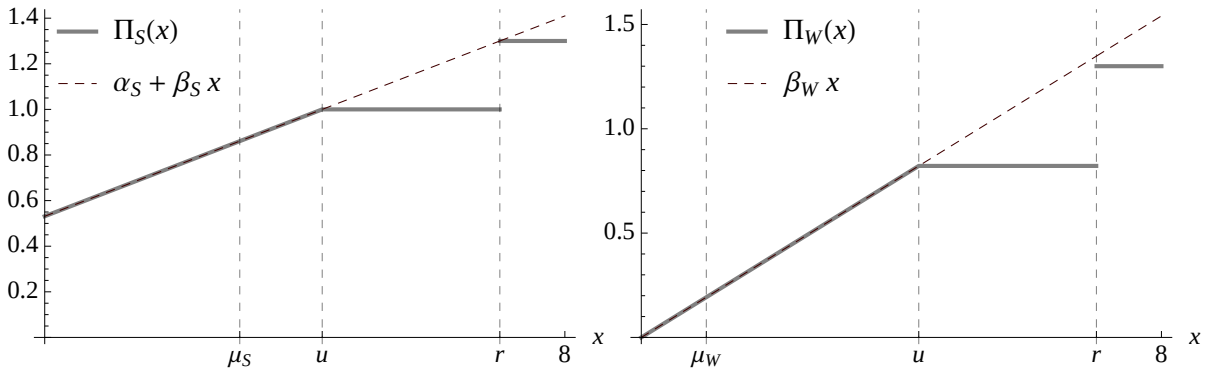
In this equilibrium, both contestants' capacity constraints bind.

In Eq1, the strong contestant chooses a uniform distribution over $[0, u]$, with probability $1-p_S^r$, and a point mass at the bonus threshold, r, with probability p_S^r; the weak contestant randomizes by choosing a point mass at 0, with probability p_W^0, and the same uniform distribution as the strong contestant, with probability $1 - p_W^0$.

Because the bonus payment, b, is positive, performance just below the bonus threshold, r, is dominated by performance equal to the bonus threshold. Thus, the upper bound of the region over which the contestants compete purely for rank dominance, u, is always less than the bonus threshold, r. Consequently, one can think of the region $(0, u]$ as the *pure rank-competition* region. In this region, the bonus is not captured and lowering performance lowers the probability of topping the rival contestant.

The distributions specified in Lemma 2.3.3 are best replies if and only if the graphs of the contestants' reward functions lie weakly below their support lines and, at all performance levels, x, in the support of their performance distributions, the support line evaluated at x equals the contest reward function evaluated at x. These optimality conditions are illustrated in Figure 2.2. Because the rank-based reward is earned if and only if the performance tops rival's performance and ties do not occur in equilibrium, the probability that a contestant wins the rank reward for a performance level of x is $F_j(x)$ where j represents the performance distribution of the rival. The contest reward functions of the contestants jump up by b at $x = r$, the point where the bonus threshold is attained.

Note that part of the strong contestant's capacity is diverted to targeting the bonus. Thus, the upper bound on the uniform distribution, u, over the pure rank competition region must be less than $2\,\mu_S$, the upper bound when all of the strong contestant's capacity is devoted to rank competition. Because $u < 2\,\mu_S$, the definition of p_W^0 provided in Lemma 2.3.3 implies that p_W^0, the probability that the weak contestant places point mass on 0 in configuration Eq1, is less than $(\mu_S - \mu_W)/\mu_S$, the probability when both contestants ignore the bonus.



A. Contestant S B. Contestant W

Figure 2.2: *Equilibrium rewards in Eq1 equilibria, when only S chases the bonus.* The figure illustrates the reward functions, Π_i and support lines, $\alpha_i + \beta_i\, x$ for contestant S (Panel A) and contestant W (Panel B). The horizontal axis represents performance, x. The parameters are $\mu_S = 3$, $\mu_W = 1$, $r = 7$ and $b = 0.3$.

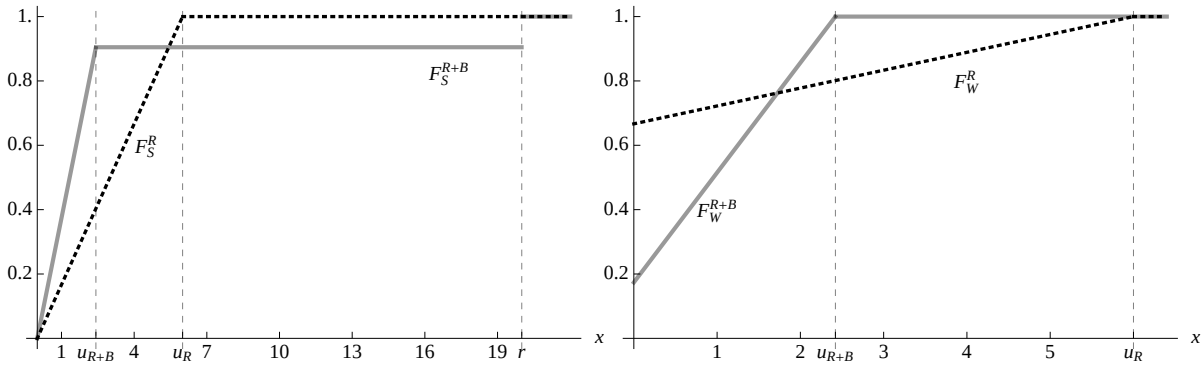
Proposition 2.3.1 (**Atalanta effect—Eq1**) *In configuration Eq1, (i.e., when only one contestant chases the bonus), the probability of downside risk taking, p_W^0 is less than the probability of downside risk taking in the absence of bonus compensation, $(\mu_S - \mu_W)/\mu_S$.*

In configuration Eq1, the reduction in downside risk taking by the weak contestant can be quite dramatic when the bonus threshold is very high and the size of the bonus is correspondingly large. This case is illustrated in Figure 2.3. When the bonus threshold is high, attaining the bonus requires a large diversion of capacity from rank competition by the strong contestant, resulting in large reduction in downside risk taking by the weak contestant. In the example in the figure, the weight placed on zero by the weak contestant is reduced from 0.67 to 0.17 by the introduction of bonus compensation.

Configuration Eq2: Both S and W chase the bonus

When the conditions for configuration Eq1 is violated, bonus chasing is tempting for the weak contestant as well as the strong contestant. When this occurs, our third configuration, Eq2, characterizes equilibrium behavior. In the Eq2 configuration, contestant performance distributions must satisfy the following conditions: (i) Each contestant's reward at the bonus threshold must meet her support line. (ii) The slope of each contestant's support line above and below the bonus threshold must be the same. (iii) For both contestants, the capacity constraint is binding. (iv) For both contestants, all points in the support of their performance distributions lie on their support lines. These conditions yield the following characterization of Eq2.

Lemma 2.3.4 (**Eq2: Both chase bonus**) *When both S and W chase the bonus, the*



A. Contestant S B. Contestant W

Figure 2.3: *Effect of bonus compensation on risk taking.* The figure presents the effect of bonus compensation equilibrium performance distributions for the strong contestant, S (Panel A), and the weak contestant, W (Panel B), in configuration Eq1, in which only S chases the bonus. In this figure the parameters $\mu_S = 3$, $\mu_W = 1$ are fixed. F_W^R (F_S^R) represents the performance distribution of W (S) in the absence of bonus competition. F_W^{R+B} (F_S^{R+B}) represents the performance distribution of W (S) when the bonus threshold $r = 20$ and bonus $b = 6$. u_R (u_{R+B}) represents the upper bound of the pure rank competition region under pure rank (rank and bonus) competition.

unique equilibrium can be characterized by the following performance distributions:

$$F_S^* = p_S^h \,\mathrm{Unif}[r, u_H] + p_S^r \,\mathbb{1}_r + (1 - p_S^h - p_S^r)\,\mathrm{Unif}[0, u_L],$$
$$F_W^* = p_W^h \,\mathrm{Unif}[r, u_H] + p_W^0 \,\mathbb{1}_0 + (1 - p_W^h - p_W^0)\,\mathrm{Unif}[0, u_L];$$

where

$$p_S^r = \frac{b\,(\mu_S - \mu_W)}{b\,r + \mu_W} \in (0,1), \qquad p_S^h = \frac{(1+b)\,(u_H - r)}{u_H} \in (0,1),$$
$$p_W^0 = \frac{(1+b)\,(\mu_S - \mu_W)}{b\,r + \mu_S} \in (0,1), \qquad p_W^h = \frac{b\,(u_H - r)}{r - u_L} \in (0,1),$$
$$u_L \in (0, r), \qquad u_H > r.$$

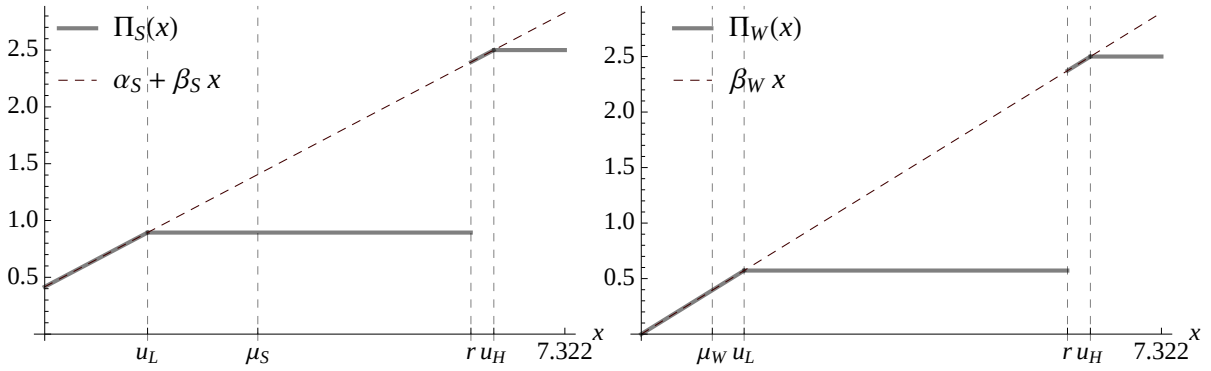
The specific functional forms of u_H and u_L are specified in Appendix Section B.1. Both contestants' capacity constraints bind.

In configuration Eq2, both contestants place positive probability weight on the region $(r, u_H]$ where performance strictly exceeds the bonus threshold. Because, over this region, sufficiently small reductions in performance always lower the probability of winning the rank reward but never affect the probability of winning the bonus reward, we term the region $(r, u_H]$, the *rank and bonus competition* region.

When performance is insufficient to capture the bonus, i.e., $x < r$, the weak contestant places probability mass p_W^0 on 0 and, with probability $1 - p_W^0 - p_W^h$, chooses a uniform

distribution over $[0, u_L]$. The strong contestant, with probability $1 - p_S^r - p_S^h$, chooses a uniform distribution over $[0, u_L]$. As in the Eq1 configuration, the weak contestant places positive weight on zero performance and the strong contestant does not and, over the pure rank competition region $(0, u_L]$, both contestants choose the same distribution of performance.

When performance is sufficient to capture the bonus, the strong manager *targets the bonus* by placing positive probability weight p_S^r exactly on the bonus threshold, while the weak contestant does not target the bonus and both contestants choose the same performance distribution over the rank and bonus competition region but target this region with different probabilities, p_S^h for the strong contestant and p_W^h for the weak contestant. Contest reward functions for this equilibrium configuration are illustrated by Figure 2.4.



A. Contestant S B. Contestant W

Figure 2.4: *Equilibrium rewards in Eq2 equilibria, when both S and W chase the bonus.* The figure illustrates the reward functions, Π_i and support lines, $\alpha_i + \beta_i\, x$ for contestant $i = S$ (Panel A) and contestant $i = W$ (Panel B). The horizontal axis represents performance, x. The parameters are $\mu_S = 3$, $\mu_W = 1$, $r = 6$ and $b = 1.5$.

The next result shows that the Atalanta effect is verified in all Eq2 configurations.

Proposition 2.3.2 (**Atalanta effect—Eq2**) *In configuration Eq2, (i.e., when both contestants chase the bonus), the probability of downside risk taking, p_W^0, is less than the probability of downside risk taking in the absence of bonus compensation, $(\mu_S - \mu_W)/\mu_S$.*

2.3.3 Upside risk taking

In this section, we will show that the effect of bonus packages on upside risk taking depends on the degree of strength asymmetry between the contestants, measured by μ_S/μ_W, the size of the bonus, and the equilibrium configuration characterizing contestant behavior.

To make these observations precise, we need to map out the regions in r–b space that sustain the three configurations, Eq0, Eq1, and Eq2. The next result shows that the three

equilibrium configurations, in fact, do partition the space of admissible bonus packages.

Lemma 2.3.5 *Any choice of admissible bonus packages, i.e., (r, b), such that $r > \mu_S$ and $b > 0$, sustains one and only one equilibrium configuration, Eq0, Eq1, or Eq2.*

In the appendix we show that, although the regions supporting the three configurations are defined by implicit polynomial equations of fairly high order, fairly simple characterizations can be obtained using parametric curves that map out the boundaries between the regions. Using this approach, the boundary between Eq0 and Eq1 in r–b space and the boundary between Eq1 and Eq2, can be represented by the image of a parametric curve, with parameter $y > 0$, as follows:

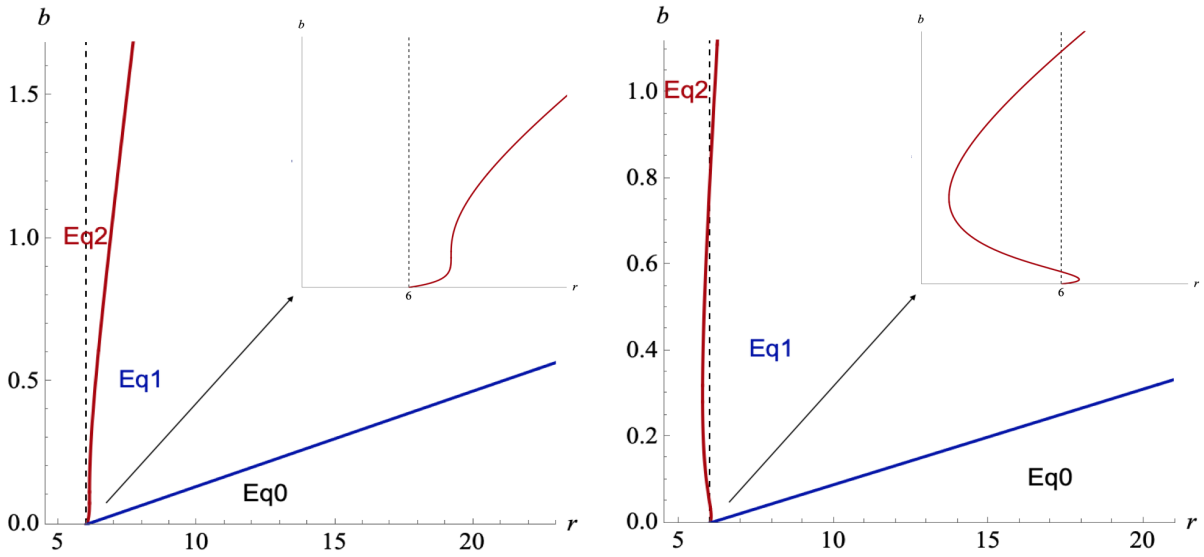
$$\mathcal{B}_0^1(y) = \frac{\sqrt{2\,y\,\mu_W + \mu_W^2} - \mu_W}{2\,\mu_S}, \qquad \mathcal{R}_0^1(y) = \frac{2\,y\,\mu_S}{\sqrt{2\,y\,\mu_W + \mu_W^2} - \mu_W};$$
$$\mathcal{B}_1^2(y) = y\left(\mu_S + \frac{y + \mu_S}{y + \mu_W}\sqrt{2\,y\,\mu_W + \mu_W^2}\right)^{-1}, \quad \mathcal{R}_1^2(y) = \mu_S + \frac{y + \mu_S}{y + \mu_W}\sqrt{2\,y\,\mu_W + \mu_W^2}.$$

The Eq1/Eq2 boundary equals the (r, b) pairs that satisfy $r = \mathcal{R}_1^2(y)$ and $b = \mathcal{B}_1^2(y)$, for some $y > 0$. The Eq0/Eq1 boundary equals the (r, b) pairs that satisfy $r = \mathcal{R}_0^1(y)$ and $b = \mathcal{B}_0^1(y)$, for some $y > 0$.

In Figure 2.5, we provide graphs of these curves. In Panel A, the ratio between the strength of the strong and weak contestant is 5:1; in Panel B, the ratio is 7.5:1. We see that, in both panels, at $y = 0$, $(r, b) = (2\,\mu_S, 0)$ and that, at $y = 0$, tangent slope of the Eq1/Eq2 boundary equals $1/(2\,\mu_S) = 1/6$. It is also easy to verify that, in both panels, the asymptotic slope of the Eq1/Eq2 boundary is given by $1/(2\,\mu_W)$. Thus, the initial behavior and asymptotic behavior of both curves are quite similar in the two panels.

However, in Panel A, where strength asymmetry is not very extreme, the curve always moves to the north-east, away from the dashed line representing the $r = 2\,\mu_S$ line. In Panel B, where strength asymmetry is very extreme, the curve bends back and transversally intersects the $r = 2\,\mu_S$ line. Thus, in Panel A, where strength asymmetry is less extreme, Eq1 configurations can be sustained only when the bonus threshold, r exceeds maximum contestant performance under pure rank competition, $2\,\mu_S$. However, in Panel B, where strength asymmetry is extreme, for some choices of the bonus level, b, Eq1 equilibria can be sustained even when at bonus thresholds, r, is less than $2\,\mu_S$.

In Eq2 equilibria where both contestants chase the bonus, upside risk taking depends on the bonus threshold, r, and length of the rank plus bonus compensation region, $u_H - r$. Bonus compensation both makes competing for rank rewards at performance levels above the bonus threshold attractive relative to competing at performance levels below the threshold



A. $\mu_S/\mu_W = 5$ B. $\mu_S/\mu_W = 7.5$

Figure 2.5: *Parametric curves for equilibrium configurations.* In the figure, the blue line represents the Eq1/Eq2 boundary and the red line represents the Eq0/Eq1 boundary. In Panel A, $(\mu_S, \mu_W) = (3, 0.6)$. In Panel B, $(\mu_S, \mu_W) = (3, 0.4)$. The dotted vertical line represents points where $r = 2\,\mu_S$, maximum performance in pure rank contests. In the top right corner of each graph, we provide a magnified view of Eq1/Eq2 boundary when b is small.

and makes competing for rank rewards less attractive than competing for bonus rewards. Given these insights the following proposition should not be too surprising.

Proposition 2.3.3 *If the equilibrium configuration is Eq1,*

(i) when strength asymmetry is not extreme, i.e., $\mu_S/\mu_W \leq 3 + 2\sqrt{2} \approx 5.83$*, maximum contestant performance,* r*, weakly exceeds maximum contestant performance under pure rank competition,* $2\,\mu_S$*;*

(ii) when strength asymmetry is extreme, there exist bonus packages, (r, b)*, under which maximum contestant performance,* r*, is less than maximum performance under pure rank competition,* $2\mu_S$*.*

If the equilibrium configuration is Eq2,

(iii) maximum contestant performance, u_H*, is increasing in the bonus payment,* b*, for* b *sufficiently small, and decreasing for* b *sufficiently large. For all bonus payments* $b > 0$*,*

$$r \leq u_H(b) \leq \min\left[\mu_S + \sqrt{(r-\mu_S)^2 + \mu_S^2}, r + \frac{\mu_S}{b}\right].$$

Thus, if b *is sufficiently large and* $r < 2\mu_S$*, i.e., maximum contestant performance is less than under pure rank competition.*

The intuition for parts (i) and (ii) of the proposition is that, if the bonus threshold is

less than maximum performance under pure rank competition, holding the strategy of the strong contestant fixed, chasing the bonus is attractive to the weak contestant. Thus, both the strong and weak contestant chase the bonus when their capacities are not too asymmetric.

If, instead, after the introduction of bonus competition, the adjustment in equilibrium strategies results in an Eq1 equilibrium, the payoff to the weak contestant from rank competition must increase sufficiently to dominate targeting a bonus with a threshold lower than the weak contestant's maximum performance under rank competition. This increase can be affected only if the strong contestant diverts significant capacity to bonus targeting and the bonus associated with the threshold is not very large. Diverting significant capacity to targeting a small bonus is only optimal for the strong contestant if the strong contestant's marginal gain from rank competition is very small. The strong contestant's marginal gain is only very small when the weak contestant is very "easy to beat," i.e., strength asymmetry is extreme.

Now consider part (iii) of Proposition 2.3.3. Bonus compensation has two effects on contestants' incentives to compete at performance levels higher than the bonus threshold: a *subsidy effect* and an *opportunity-cost effect*. The bonus payment, b, provides a subsidy for rank competition above the bonus threshold but not for rank competition below the bonus threshold. This effect favors an increase in the length of the rank and bonus competition region relative to the pure rank competition region and, consequently, an increase in the upper bound of the rank and bonus competition region, u_H.

At the same time, bonus packages generate an opportunity cost for rank competition at performance levels above r: capacity devoted to rank competition could have been devoted to targeting the bonus. *Ceteris paribus*, the opportunity-cost effect leads contestants to place more weight on bonus targeting relative to rank competition and this effect favors a reduction of the length of the rank and bonus competition region.[13]

Figure 2.6 illustrates the interaction between the subsidy and opportunity-cost effects by plotting maximum performance, u_H, as a function of the level of bonus compensation, b. The subsidy effect dominates at low bonus levels while the opportunity cost effect dominates at high bonus levels.

[13]In general, the relation between the upper bound of performance in Eq2 equilibria, u_H, and the level of bonus compensation is not always quasiconcave (examples provided upon request) and, in general $b \hookrightarrow u_H(b)$ can cross the upper bound of of performance under pure rank competition, $2\,\mu_S$, more than once. However, unless the bonus threshold, r, is very close to $2\,\mu_S$ *and* strength asymmetry is also fairly close to the extreme asymmetry boundary, $3+2\sqrt{2}$, u_H will cross $2\,\mu_H$ once from above. The statement of the exact conditions required for a single crossing, which involves radical solutions to a cubic polynomial, is a bit cumbersome and thus is relegated to Lemma B.1.1 in Appendix B.1.

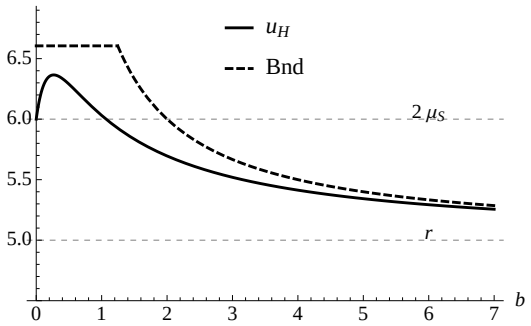


Figure 2.6: *Subsidy effect vs. opportunity-cost effect.* In the example, $\mu_S = 3$, $\mu_W = 2$, and $r = 5$. $2\,\mu_S$ represents the upper bound on performance under pure rank competition. Bnd $= \min\left[\,\mu_S + \sqrt{(r-\mu_S)^2 + \mu_S^2}, r + \mu_S/b\,\right]$ represents the upper bound on u_H defined in Proposition 2.3.3.(iii).

2.3.4 How low can downside risk taking go?

Figure 2.3 in Section 2.3.2 illustrates that bonus compensation can reduce downside risk taking dramatically. This example naturally raises the question of whether bonus compensations can eliminate downside risk taking entirely. As the next proposition shows, in fact, in all equilibria, the probability of downside risk taking is positive. However bonus packages exist which support equilibria in which the probability of downside risk taking is arbitrarily small. in the limit, these packages, can be very "expensive" in terms of the expected bonus compensation received by the strong contestant.

Proposition 2.3.4

(i) There exists no equilibrium in which downside risk taking is absent (i.e., $p_W^0 = 0$).

(ii) However, there exists a sequence of bonus packages, $\{(r_n, b_n)\}$ which

- *(a) sustains Eq1 equilibria;*
- *(b) satisfies $r_n \to \infty$, $b_n \to \infty$, $\frac{b_n}{r_n} \to \frac{1}{2\,\mu_W}$ as $n \to \infty$;*
- *(c) yields a downside risk taking approaching 0, i.e., $p_W^0(n) \to 0$ as $n \to \infty$.*
- *(d) In the limit, the distributions of contestant performance approach the equilibrium distribution in a pure rank competition in which both contestants' capacities equal μ_W.*

(iii) The limiting expected value of bonus compensation under these bonus packages equals $\frac{1}{2}\,\frac{\mu_S - \mu_W}{\mu_W}$.

Thus, a limiting sequence of bonus compensation packages, $\{(r_n, b_n)\}$ does exist which reduces downside risk taking to an arbitrarily large extent. Under this sequence of bonus packages, the expected bonus payment to the strong contestant, $p_S^r \times b$ converges to $(\mu_S - \mu_W)/\mu_W$. This payment can be quite large. For example if $\mu_W = 1/2$, the asymptotic

bonus payment equals $\mu_S - \mu_W$, the entire difference in capacity between the strong and weak contestants.

2.3.5 Bonus compensation and overall risk?

The analysis thus far has shown that bonus compensation can reduce downside risk taking. However, bonus compensation has other effects on the distributions of contestants' performance, e.g., it sometimes increases upside risk taking. In some cases, the weak contestant's distribution under rank and bonus competition second-order stochastically dominates the weak contestant's distribution under pure rank competition. In fact, this case is illustrated in Figure 2.3. However, in general, the performance distributions engendered by bonus packages are neither second-order stochastically dominant or dominated by the performance distributions engendered by pure rank competitions. Thus, the performance distributions engendered by bonus compensation will be preferred by some risk averse agents but not by others.

We aim to examine the question of whether the effects on risk taking produced by bonus compensation will be valued by a risk-averse agent endowed with a "reasonable" utility-of-consumption function. For this reason, we choose the utility specification most often used in field studies and laboratory experiments that calibrate risk aversion—constant relative risk aversion (CRRA). We parameterize the CRRA function by assuming a risk aversion coefficient of 3, a coefficient within the range of estimated risk aversion coefficient in a number of studies (Chiappori and Paiella, 2011; Garcia et al., 2003; Mankiw, 1985; Friend and Blume, 1975). In order to ensure that our results are not produced by unreasonably large, in expected value terms, bonus payments, we assume that the risk-averse agent's consumption is reduced by bonus payments to the contestants. The utility of a risk-averse agent is thus determined by endowed wealth and the sum of two contestants' performance less the bonus payments received by the contestants.

In our example $\mu_S = 3$ and $\mu_W = 1$, and the risk-averse agent's consumption endowment equals 0.50.[14] Using a numerical global optimization program, we determine the bonus package that maximizes the utility of the risk-averse agent given the equilibrium strategies of the contestants. The results of this exercise are reported in Table 2.1.

As we see from the table, the utility maximizing bonus package supports an Eq1 equilibrium configuration. At the maximizing bonus package, the weak contestant submits 0 performance with probability p_W^0 and otherwise randomizes uniformly over the pure rank

[14]Given the equilibrium distribution of contestants' strategies, a positive endowment is required for expected utility of performance to be defined for CRRA utility-of-consumption functions with conventional risk-aversion coefficients.

Eq Config.	*Optimal package*		*Contest outcome*			*Designer payoff*	
						Optimal	Pure
	r^*	b^*	u	p^0_W	p^r_S	package	rank
Eq1	6.467	0.156	4.726	0.577	0.155	-32.475	-33.424

Table 2.1: *Optimal bonus compensation package.* In this example, the risk-averse agent's utility of consumption is given by $u(w) = w^{1-\theta}/(1-\theta)$, $\theta = 3$. The capacities of the strong and weak contestants are given by $\mu_S = 3$ and $\mu_W = 1$, respectively. Consumption consists of an endowment equal to 0.5 and the sum of the performance of the two contestants less the cost of the bonus payment.

competition region $(0, u]$. With probability $1 - p^r_W$, the strong contestant also randomizes uniformly over the pure rank competition region, and, with probability p^r_S chooses performance exactly equal to the bonus threshold, r.

Bonus compensation is positive, i.e., the risk-averse agent's utility is increased by the provision of bonus compensation, even though bonus payments reduces the agent's consumption. The expected cost of optimal bonus compensation, $p^r_S \times b \approx 0.024$, is quite modest relative to total expected performance, $\mu_S + \mu_W = 4$. Not surprisingly, the package is also far less costly than the limit of the downside risk elimination sequence, whose limiting expected value, as shown in Proposition 2.3.4, equals 1. Nevertheless, the reduction in downside risk taking is significant, from $0.6\bar{6}$, in the absence of bonus compensation (see Section 2.2.2), to 0.577, under the utility-maximizing bonus package. This example shows, for conventional specifications of risk aversion, offering bonus compensation can increase the certainty-equivalent value of contestant performance net of bonus payments. *A fortiori*, this implies that the certainty-equivalent value of performance gross of bonus compensation is higher when bonus compensation is offered.

2.4 Robustness of the Single-Period Baseline Model

The intuition we provided for our results was that, because strong contestants' marginal rewards from rank competition are smaller than weak contestants, they are more likely to be distracted by bonus compensation. Distraction makes strong contestants weaker rank competitors and thus reduces the amount of downside risk weak competitors need to assume to compete with strong competitors. Thus, our intuition rests on two conditions, whose satisfaction is verified in the baseline model: (a) In pure rank competitions, the strong contestants' marginal reward from capacity invested in rank competition is less than

the weak contestants' marginal reward. (b) Diverting capacity to bonus chasing makes strong contestants weaker rank competitors.

Condition (a) relates to our specification of pure rank competition. The only obvious way to generalize our pure rank competition model within the risk-taking contest framework is to increase the number of contestant. In Section 2.4.1 we argue that condition (a) is robust to this generalization. In Section 2.4.2 we show that, under linear absolute performance rewards, e.g., pure equity compensation for the contestants, property (b) is not satisfied because contestant strategies are not affected by linear rewards. In Section 2.4.3 we show that when bonus packages are not challenging, property (b) is sometimes not satisfied because bonus chasing need not make the strong contestants weaker rank competitors.

2.4.1 Multiple contestants

In Appendix Section B.2, we show that in contests with more than two contestants, condition (a) is always satisfied when no two contestants have the same capacity. Thus, there is no reason to suspect that simply adding more contestants lead to different qualitative conclusions than the baseline model. Solving the general problem of characterizing equilibrium behavior in all asymmetric multi-contestant rank-and-bonus risk taking contests is a worthy objective. But is not a trivial task and not the objective of this paper. In fact, characterizing all possible equilibrium configurations in pure-rank, three contestant all-pay auctions which, as discussed in the introduction, are closely related to risk taking contests, is quite a challenging task that has only recently been accomplished (Xiao, 2018). However, as we illustrate by an example in Appendix Section B.2, it is easy to numerically construct multi-contestant contest equilibria in which the Atalanta effect is verified.

2.4.2 Linear compensation

In order for condition (b) to hold, it clearly must be the case that compensation packages affect contestant behavior. However, contestants receiving linear compensation will play the same strategies as they would in a pure rank competition contest. To see this, let γ represent the linear compensation coefficient. Contestant payoffs in the linear compensation setting would still be characterized by the support line conditions provided by Remark 2.3.3. If we let α', β' represent the multipliers for the support lines under linear compensation, the equilibrium best-reply conditions can be expressed as

$$\begin{aligned} \gamma\, x + F_W(x) &= \alpha'_S + \beta'_S\, x,\ x \in \mathrm{Supp}_S, && \text{for all } x \geq 0,\ x + F_W(x) \leq \alpha'_S + \beta'_S\, x; \\ \gamma\, x + F_S(x) &= \alpha'_W + \beta'_W\, x,\ x \in \mathrm{Supp}_W, && \text{for all } x \geq 0,\ x + F_S(x) \leq \alpha'_W + \beta'_W\, x. \end{aligned} \tag{2.4.4}$$

If we let $\alpha_i' = \alpha_i$ and $\beta_i' = \beta_i + \gamma$, $i = S, W$, where α_i and β_i represent the multipliers for the Eq0 equilibrium, we see that the distribution functions specified in Lemma 2.3.1 satisfy the equilibrium conditions given by equation (2.4.4). Thus, the Eq0 equilibrium distribution functions are equilibrium distribution functions under linear compensation.

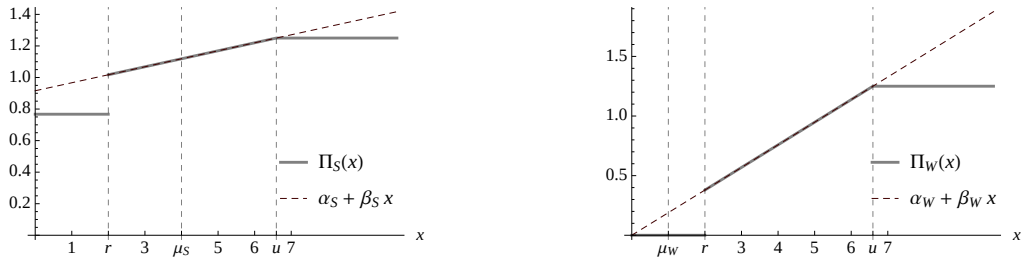
Most other features of the compensation contract assumed in our analysis—e.g., a fixed bonus payment for reaching a specific target level of performance—are less important for verifying the Atalanta effect. However, the boundedness of bonus compensation does play an important role. The problem with unbounded compensation contracts is not that such contract reverse the Atalanta effect. Rather, when compensation is unbounded, equilibria may not exist. The non-existence problem could be surmounted by either imposing sufficient contestant risk aversion or by judiciously imposing a cap on performance. Resorting to these devices would simply result in a less tractable model without producing new insights.

2.4.3 Easy bonus compensation

When bonus packages are *easy*, i.e., not challenging, their attainment may not compel strong contestants to sacrifice rank dominance. Thus, condition (b) may not hold. Easy bonus packages are packages impose bonus targets under which at least one contestant is able to attain the target with certainty. Consider the case where only the strong contestant can attain the bonus with certainty, i.e., the case where the bonus threshold satisfies $\mu_W < r < \mu_S$. For the sake of brevity we consider specific examples, more general algebraic characterizations of these equilibrium configurations are provided in Appendix Section B.5.

Example 2.4.1 (Low bonus payment, b) *In this example, $\mu_S = 4$, $\mu_W = 1$, $r = 2$, and $b = 0.25$. Because $\mu_S > r$, this case represents an easy bonus package. Using the same approach to computing equilibria as used for challenging bonus packages in previous sections yields the following equilibrium: the weak contestant randomizes between a performance of 0, chosen with probability $p_W^0 = {}^{23}/_{30}$ and a uniform performance distribution over $[r, u]$, where $u \approx 6.6$. The strong contestant places mass on the bonus threshold r, with probability $p_S^r \approx 0.13$, and selects the same uniform performance distribution with probability $1 - p_S^r$. The reward functions, Π, for the strong and weak contestant are graphed in Figure 2.7.*

In this equilibrium, the strong contestant, S, skips the pure rank competition region entirely and concentrates all performance at levels at or above the bonus threshold. This is possible because the bonus package is easy, i.e., features a bonus threshold lower than S's capacity. For this reason, in contrast to the challenging bonus packages case, chasing the bonus does not make the strong contestant a weaker competitor over the pure rank competition region.



A. Contestant S B. Contestant W

Figure 2.7: *Equilibrium reward functions under an easy bonus package: Small bonus.* The figure illustrates the reward functions, Π_i and support lines, $\alpha_i + \beta_i\, x$ for contestant S (Panel A) and contestant W (Panel B). The horizontal axis represents performance, x. The parameters are $\mu_S = 4$, $\mu_W = 1$, $r = 2$ and $b = 0.25$.

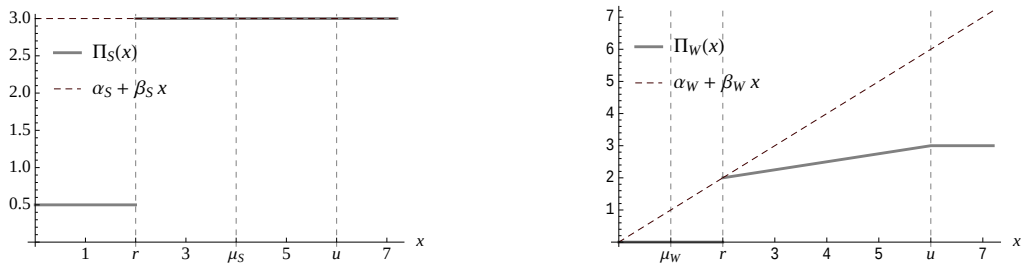
In fact, because the strong contestant always submits performance higher than the bonus threshold, any performance in the pure rank competition region by the weak contestant is sure to result in losing the rank competition.

In this example, the bonus payment is fairly small relative to the rank reward, 1. Thus, the weak contestant W is still motivated to compete with S for rank rewards. Because, S does not compete over the pure rank competition region, to compete for rank, W must also submit performance in excess of the bonus threshold, r. Given W's low capacity, this requires assuming considerable downside risk. In fact, the probability of downside risk taking in this equilibrium equals $p^0_W = {}^{23}/_{30} > {}^{3}/_{4}$, the probability of downside risk taking in a pure rank competition. Consequently, in this example, easy bonus compensation increases downside risk taking.

Now consider another example, fix all the parameters as in the first example except increase b from 0.25 to 2. This increase in the bonus level is sufficient to lead the weak contestant to give up on rank-competition entirely and reverses the conclusions of Example 2.4.1.

Example 2.4.2 (High bonus payment, b) *Suppose that $\mu_S = 4$, $\mu_W = 1$, $r = 2$, and $b = 2$. The equilibrium in this example calls for W to place mass on 0 with probability $p^0_W = {}^{1}/_{2}$ and place mass on the bonus threshold, with probability ${}^{1}/_{2}$; S uniformly randomizes over $[r, u]$, where $u = 6$. Equilibrium reward functions for the two contestants are illustrated in Figure 2.8.*

In contrast to Example 2.4.1, the gains to W from attempting to compete for rank with a much stronger rival are not as attractive as chasing the very large bonus payment. Thus, W skips rank competition entirely, and adopts a pure bonus chasing strategy, conceding rank dominance to S. The fact that bonus targeting is more attractive to W than rank competition with S implies that the amount of downside risk that W must accept to compete with S for rank dominance is less than the amount of downside risk W must accept to target



A. Contestant S B. Contestant W

Figure 2.8: *Equilibrium reward functions under an easy bonus package: Large bonus.* The figure illustrates the reward functions, Π_i and support lines, $\alpha_i + \beta_i\, x$ for contestant S (Panel A) and contestant W (Panel B). The horizontal axis represents performance, x. The parameters are $\mu_S = 4$, $\mu_W = 1$, $r = 2$ and $b = 2$.

the bonus payment. Thus, the equilibrium probability of downside risk taking $p^0_W = 1/2$ *is less than the probability of downside risk taking under pure rank competition,* $3/4$.

Examples 2.4.1 and 2.4.2 show that introducing easy bonus packages into rank competitions can either increase or decrease downside risk taking. Because the strong contestant eschews pure rank competition, chasing the bonus does not make the strong contestant a weaker competitor. Although bonus compensation can reduce downside risk taking (Example 2.4.2) the mechanism by which bonus compensation affects risk-taking is very different from the mechanism driving the Atalanta effect.[15]

As we can see, the effects of easy bonus packages on risk taking are quite different than the effect of challenging bonus packages. This raises the question of why we focus our analysis on challenging packages. Our focus is motivated by three considerations. First "challenging," at least as we define this term, appears to be a realistic description of most actual bonus packages. Even CEOs, whose bonus targets are notoriously unambitious, miss targets about 40% of the time (Kay et al., 2015). Unless the probability of meeting challenging targets is close to zero, this implies that between half and all of CEO bonus packages are challenging, i.e., there is some positive probability that bonus targets cannot be attained.

Second, easy bonus packages, do not, in the absence of rank competition, encourage risk-taking and are not the sort of compensation packages analyzed in the literature on bonus compensation and risk taking. Third, while the Atalanta effect might be viewed as being

[15]By making bonus rewards even easier, i.e., setting $r < \mu_W$, and offering a sufficiently large bonus payment, it is also possible to support equilibria in which both S and W submit performance at or above the bonus threshold. In such equilibria, bonus payments, in equilibrium, are fixed non-random rewards conditioned on not submitting performance below the bonus threshold. The lower bound for net performance, performance less the fixed payment, is zero, and the probability of net performance equaling 0 is higher than the probability of performance equaling 0 under pure rank competition. See Appendix Section B.5 for details.

surprising—challenging bonus packages, which, in the absence of rank rewards, encourage risk taking, discourage downside risk taking by rank-motivated contestants—it is not particularly surprising that easy bonus packages, which do not encourage risk taking in the absence of rank competition, may not encourage risk-taking by rank-motivated contestants.

2.5 Dynamic Contests

Risk taking in rank competitions results from asymmetric competitor strength. Firms, banks, mutual funds, organizers of sporting contests, typically do not aim to set competitions between competitors with wildly differing abilities. Thus, in practice, risk taking appears to be the result of a contestant, say a trader, suffering an adverse shock, e.g., a large trading loss, and thereby "falling behind." The trailing contestant, realizing that without taking big risks, he has little or no chance of besting his rivals, tries to catch up by "doubling down," i.e., through employing high risk strategies. The fact that risk taking in rank contests appears, in practice, to typically result from the evolution of dynamic contests naturally raises the question of whether the Atalanta effect persists in a dynamic contest, or, if the new incentives introduced by dynamic competition are sufficient to nullify the effect.

To address this question, we consider the Atalanta effect in the context of dynamic contests where contestants, *ex-ante*, have equal capacity. We first develop, in Section 2.5.1, the general framework for our dynamic predation contest. Next, in Section 2.5.2 we parametrically specify the general formulation and solve for equilibrium strategies.

2.5.1 Dynamic contests–general formulation

We incorporate dynamics by assuming an infinite date world with time indexed by t. In each period, contestants compete for both rank and bonus-based period rewards. Performance in the current period affects capacity for performance in the next period. In this dynamic setting, the capacity of contestants evolves over time based on past performance. Thus, we cannot follow the approach used in the previous section of labeling the contestants "strong" and "weak." Instead, we label the contestants F and G, and apply the same labels to their performance distributions.

Initially, the two contestants have identical capacities equal to $\mu > 0$. If, in period t, the capacity of a given contestant equals μ_t, then, if the contestant wins the period-t rank competition, the contestant's capacity in period $t+1$ increases to $\mu_t + \delta$, where $\delta = \mu/m$, for some $m \in \{2, 3, 4 \ldots\}$. If the contestant loses the rank competition in period t, the

contestant's capacity in period $t+1$ falls to $\mu_t - \delta$. If a contestant's capacity falls to zero, she becomes insolvent and is "knocked out," i.e., cannot compete in future rounds. In this event, the surviving contestant has no competitor. Thus, surviving contestant's capacity no longer changes and the survivor always wins the rank reward. Because the potential gains for the contestants are unbounded in this infinite date setting, we introduce a discount factor ϕ, where $\phi \in (0,1)$, which is applied to continuation values.

We restrict attention to symmetric equilibria in which contestants play symmetric Markovian strategies. The assumption of Markovian strategies implies that strategies depend only on capacities at the start of the given period, μ_F and μ_G. Thus strategies are a function of the vector, (μ_1, μ_2), where the first component of the vector equals the capacity of F and the second the capacity of G. These vectors are the states of the Markovian system, we represent state with ω and represent the set of all states by Ω.

The fact that contestants play symmetric strategies implies that the strategy played by G when $\mu_F = \mu_1$ and $\mu_G = \mu_2$ is the same as the strategy played by G when $\mu_F = \mu_2$ and $\mu_G = \mu_1$. Recalling that the first component of the state vector is the capacity of F, we see that the strategy played by G when the state is (μ_1, μ_2) equals the strategy played by F when the state is (μ_2, μ_1). Thus, to simplify notation, define, for any state $\omega = (\mu_1, \mu_2)$, the *reversed state* $\tilde{\omega}$ by $\tilde{\omega} = \widetilde{(\mu_1, \mu_2)} = (\mu_2, \mu_1)$. Using this definition, the symmetry assumption can be expressed simply as the requirement that, the equilibrium strategy for contestant G in state ω equals the equilibrium strategy for contestant F in state $\tilde{\omega}$. Thus, we need only solve for the equilibrium strategy of contestant F.

A strategy for contestant F in state ω is a performance distribution. At each state, F's performance distribution is restricted by the capacity constraint, where capacity is determined by the state. Thus, at state ω, the set of feasible performance distribution for F is given by

$$\mathcal{F}_\omega = \mathcal{F}_{(\mu_F, \mu_G)} = \Big\{ F : F[0-] = 0 \text{ and } \int x \, \mathrm{d}F(x) \leq \mu_F \Big\}. \tag{2.5.5}$$

Thus, the value function for F will, for each state vector ω map the current capacity vector, (μ_F, μ_G) into value, i.e., $V_F : \Omega \to \mathbb{R}$, where value is defined to equal the expected discounted rewards in the period and all subsequent periods. The dynamics of the value functions are described by a Markovian process in which the current state of the system is ω. The states $(0, 2\,\mu)$ and $(2\,\mu, 0)$ are the absorbing states for the system. The set of these states is represented by $\mathcal{A}$. Because a best response for either F or G will never lead to a zero probability of winning the period rank contest, the set of $2\,m+1$ recurrent states,

represented by $\mathcal{NA}$, is given by

$$\mathcal{NA} = \left\{ \left(\mu \left(1 + \tfrac{i}{m}\right), \mu \left(1 - \tfrac{i}{m}\right)\right) : i \in \{1 - m, 2 - m, \ldots 0 \ldots m - 2, m - 1\} \right\}.$$

If, at a recurrent state, one contestant's capacity is greater than the other's, we will term the contestant with greater capacity the *leading* contestant and the contestant with less capacity the *trailing* contestant.

Computing values at the absorbing states, $\mathcal{A}$, is simple. If G's capacity is 0, and thus G has been knocked out of the competition, F is sure to receive the rank reward, 1. Thus, F's optimal performance distribution in the current period and thereafter maximizes the expected bonus reward. Given that capacity equals $2\,\mu$ implies that F's period payoff equals $1 + \frac{2\,\mu\,b}{r}$ in each period. If F has been knocked out, then clearly F's payoff is 0 in the current and all subsequent periods. Discounting these period payoffs yields the value that F must receive at the absorbing states in any equilibrium:

$$V_F(2\,\mu, 0) = \frac{2\,b\,\mu + r}{r\,(1 - \phi)}, \quad V_F(0, 2\,\mu) = 0.$$

Now consider the payoff for contestant F in a recurrent state, $\omega \in \mathcal{NA}$. F will choose a random performance, $X \overset{d}{\sim} F$ in the current period. This will affect F's current period reward, determined by the structure of bonus compensation and G's performance distribution. In addition, performance in the current period will affect the probability of state transitions: if the contestant tops her rival in the current period, i.e., $X_F > X_G$, her capacity will increase by δ while her rival's capacity will decrease by δ. Similarly, if $X_F < X_G$ then her capacity will decrease by δ and her rival's will increase by δ. The effect of state transitions on F's payoffs will depend on the value function, V_F, assumed.

To capture these effects, define, for $\omega = (\mu_F, \mu_G) \in \mathcal{NA}$, the function $\mathcal{T}$ by

$$\begin{aligned}\mathcal{T}[V_F, F, G](\mu_F, \mu_G) = &\int \left(G(x) + b\,\mathbb{1}_r(x)\right) \mathrm{d}F(x) + \\ &\phi\left(\mathbb{P}(X_F > X_G)\,V_F(\mu_F + \delta, \mu_G - \delta) + \mathbb{P}(X_F < X_G)\,V_F(\mu_F - \delta, \mu_G + \delta)\right). \quad (2.5.6)\end{aligned}$$

Note that $\mathcal{T}$, maps value functions, V_F, and current period contestant performance distributions, F and G, into a possibly different value function given by $\mathcal{T}[V_F, F, G](\mu_F, \mu_G)$. An equilibrium value function satisfies the consistency or "promise keeping" condition of dynamic programming—the value function produced by $\mathcal{T}$ equals the value function input into $\mathcal{T}$. At the same time, the equilibrium performance distribution selected in each state must also be a best response for each contestant, given the equilibrium value function and the performance distribution selected by the rival, G. These conditions, combined with

symmetry condition (which implies that $G_\omega = F_{\tilde{\omega}}$), yields our definition of an equilibrium in the dynamic contest game:

Definition 2.5.1 *An equilibrium is a collection of distributions* $\{F_\omega : \omega \in \mathcal{NA}\}$*, and a function* $V^* : \Omega \to \mathbb{R}$ *satisfying the following conditions:*

1. *For all* $\omega \in \mathcal{NA}$*,*

$$V_F^*(\omega) = \max\{\mathcal{T}[V_F^*, F, F_{\tilde{\omega}}^*](\omega) : F \in \mathcal{F}_\omega\} \text{ and} \tag{NA:PK}$$

$$\mathcal{T}[V_F^*, F_\omega, F_{\tilde{\omega}}^*](\omega) = \max\{\mathcal{T}[V_F^*, F, F_{\tilde{\omega}}^*](\omega) : F \in \mathcal{F}_\omega\}. \tag{NA:BR}$$

2. *For all* $\omega \in \mathcal{A} = \{(2\,\mu, 0), (0, 2\,\mu)\}$*,*

$$V_F(2\,\mu, 0) = \frac{2\,\mu\,b + r}{r\,(1-\phi)} \text{ and } V_F(0, 2\,\mu) = 0. \tag{A:V}$$

We solve for equilibria numerically by using the standard iteration approach to solving dynamic programming problems. The procedure is detailed in Appendix Section B.3.2. It essentially involves the following steps: First, we develop a programming problem, whose solution for any fixed value function satisfying equation (A:V), yields performance distributions for each recurrent state, $\omega \in \mathcal{NA}$ which satisfy the best response condition given by equation (NA:BR). We then use the program to define inductively a sequence of value functions (V^n), for $\omega \in \mathcal{NA}$ by

$$V^n(\omega) = \mathcal{T}[V_F^{n-1}, F_\omega^{n-1}, F_{\tilde{\omega}}^{n-1}](\omega), \quad \omega \in \mathcal{NA},$$

and, for the absorbing states $\omega \in \mathcal{A}$, we define $V^n(\omega)$ using equation (A:V). Using the contraction mapping theorem of dynamic programming, we show that $V^n \to V^*$ and that V^* satisfies the conditions of Definition 2.5.1 and thus is an equilibrium.

2.5.2 A dynamic predation game

We numerically solve the dynamic predation game in the case where $\mu = 1$ and $\delta = 1/2$. These assumptions imply that the system states, Ω, are given by $(1, 1)$, $(3/2, 1/2)$, $(1/2, 3/2)$, $(2, 0)$, and $(0, 2)$. The transition dynamics between the states is illustrated in Figure 2.9.

Because the state transition function depicted in Figure 2.9 implies that if a contestant's capacity reaches 0, the contestant is knocked out of the competition and the rival contestant thus always captures the rank reward in subsequent periods, the contest is a predation contest in which predation requires rank dominance. Thus, the continuation values for the two contestants, which will evolve solely based on success in rank competition, greatly

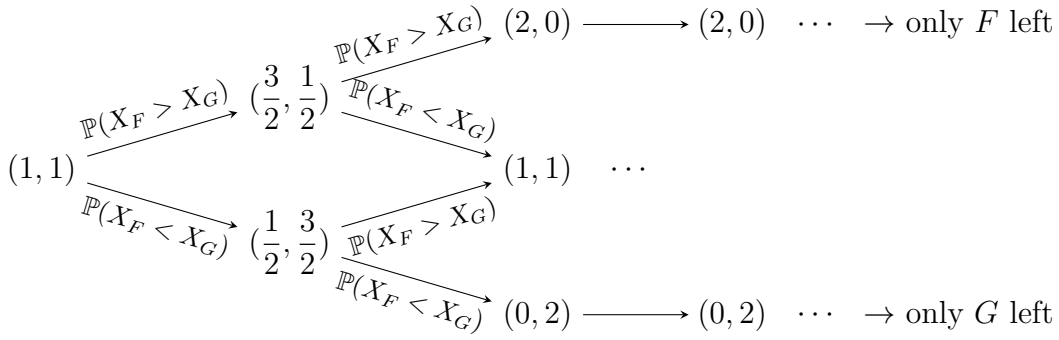


Figure 2.9: *State transition dynamics.*

augment the gains from rank competition versus bonus chasing. One might conjecture that, in this predation setting, leading contestant should concentrate on finishing off the trailing contestant. However, the following result example shows this need not be the case.

In this example, $b = 1$, $r = 3$, and $\phi = 0.9$. For the absorbing states, $(0, 2)$ and $(2, 0)$, equilibrium value function is determined by equation (A:V) of Definition 2.5.1. Table 2.2 presents the outcomes of the dynamic predation model (Panel A) and compares them with outcomes of the single-period model with the same bonus and bonus threshold (Panel B) and pure-rank dynamic predation contest (Panel C), i.e. a dynamic predation contest in which the bonus is set to 0. We only provide results at the states where contestant F is leading, as the values when G is leading are given by symmetry.

At state $(1, 1)$ neither contestant chases the bonus in the dynamic predation contest while both contestants chase the bonus in the single-period model. A greater focus on rank competition in the dynamic contest results from the rank-dependent effect of continuation values. Rank dominance increases capacity in the next period and thus increases F's probability of winning at the continuation state $(3/2, 1/2)$. If F wins again at $(3/2, 1/2)$, G is knocked out of the competition and thus F will henceforth monopolize rank rewards. The large continuation reward associated with winning the rank competition at $(1, 1)$ is evidenced by the fact that the value of F at $(3/2, 1/2)$ is approximately 10 times higher than G's value. Thus, the predation model clearly increases the focus of rank-based rewards.

At state $(3/2, 1/2)$, despite the fact that another ranking victory by F will knock G out of the competition, F diverts some capacity to bonus chasing, although much less than in the single-period model. The diversion of capacity produces the Atalanta effect—the probability that G's performance equals 0 is 0.612 in the dynamic predation contest while under dynamic pure rank competition it equals 0.667. Thus bonus payments reduce downside risk-taking.

The reduction in G's downside risk-taking is caused by F's diversion of capacity to bonus chasing. Bonus chasing by F at $(3/2, 1/2)$, when F is leading, is motivated by two factors:

	Regions				Values
	downside risk taking	pure rank competition	bonus targeting	rank+bonus competition	$(V_L^*\ ,\ V_T^*)$
A: Dynamic rank+bonus					
$(\mu_L, \mu_T) = (1, 1)$	$\{0\}$	$(0, 2]$	$\{3\}$	$\varnothing$	
$(\text{Prob}(L), \text{Prob}(T))$	(0, 0)	(1, 1)	(0, 0)	(0, 0)	(7.710, 7.710)
$(\mu_L, \mu_T) = (3/2, 1/2)$	$\{0\}$	$(0, 2.296]$	$\{3\}$	$(3, 3.062]$	
$(\text{Prob}(L), \text{Prob}(T))$	(0, 0.612)	(0.813, 0.360)	(0.123, 0)	(0.064, 0.028)	(14.614, 1.406)
$(\mu_L, \mu_T) = (2, 0)$	$\{0\}$	$\varnothing$	$\{3\}$	$\varnothing$	
$(\text{Prob}(L), \text{Prob}(T))$	(0.333, n.a.)	(0, n.a.)	(0.667, n.a.)	(0, n.a.)	(16.667, 0.0)
B: Single period rank+bonus					
$(\mu_S, \mu_W) = (1, 1)$	$\{0\}$	$(0, 1.4]$	$\{3\}$	$(3, 3.2]$	
$(\text{Prob}(S), \text{Prob}(W))$	(0, 0)	(0.875, 0.875)	(0, 0)	(0.125, 0.125)	(0.625, 0.625)
$(\mu_S, \mu_W) = (3/2, 1/2)$	$\{0\}$	$(0, 0.954]$	$\{3\}$	$(3, 3.184]$	
$(\text{Prob}(S), \text{Prob}(W))$	(0, 0.445)	(0.599, 0.465)	(0.284, 0)	(0.117, 0.090)	(1.176, 0.314)
C: Dynamic pure rank					
$(\mu_L, \mu_T) = (1, 1)$	$\{0\}$	$(0, 2]$	$\varnothing$	$\varnothing$	
$(\text{Prob}(L), \text{Prob}(T))$	(0, 0)	(1, 1)	(0, 0)	(0, 0)	(5.0, 5.0)
$(\mu_L, \mu_T) = (3/2, 1/2)$	$\{0\}$	$(0, 3]$	$\varnothing$	$\varnothing$	
$(\text{Prob}(L), \text{Prob}(T))$	(0, 0.667)	(1, 1)	(0, 0)	(0, 0)	(9.083, 0.917)

Table 2.2: *Equilibria at different states for dynamic, static, and pure rank competition.* The table presents the equilibrium strategies of the leading and trailing contestants, and compares the single-period vs. dynamic, and rank+bonus vs. pure rank cases. In the table, $r = 3$ and $b = 1$. "Prob" represents the probability that the performance of the contestants lies in a given region. (μ_L, μ_T) represent the capacities of leading and trailing contestants in the different states of the dynamic game. (μ_S, μ_W) represent the endowed capacities in the single-period game. Equilibrium values are the value pair (V_L^*, V_T^*).

First, at $(3/2, 1/2)$, for the same reasons as developed in Section 2.3.2, F's capacity advantage reduces the marginal gain to F from rank competition with her weakened rival, G. Second, the value increase produced by winning at $(3/2, 1/2)$ is much less than from winning at $(1, 1)$: knocking out G by winning at $(3/2, 1/2)$ only increases F's value from 14.614 to 16.667. At $(3/2, 1/2)$, it is very likely that F will eventually knock out G and F's value at $(3/2, 1/2)$ already impounds this fact. If F loses at $(3/2, 1/2)$, F's value will fall to 7.710. Thus, the gap between the wining and losing values is less at $(3/2, 1/2)$ than it is at $(1, 1)$. Moreover, given F's capacity advantage at $(3/2, 1/2)$, F is much more likely to win at $(3/2, 1/2)$ even if F diverts significant capacity to chasing the bonus. Thus, F's value function weighs the

small increase in value from winning more than the large decrease from losing.

Thus, when one contestant has a big lead in the dynamic contest, the trailing contestant's optimal strategy involves significant downside risk taking. At the same time, the big lead reduces the leading contestants marginal gain from focusing on rank competition. This reduced marginal gain encourages the strong contestant to chase the bonus. Bonus chasing in the dynamic contest, for the same reasons as advanced for the single-period contest, mitigates downside risk taking by the weak contestant.

2.6 Conclusion and Implication

In this paper, we analyzed the interaction between rank-based rewards, and rewards conditioned on attaining an absolute level of performance. We showed that introducing bonus compensation into rank-competitions always reduces downside risk taking and sometimes also lowers upside risk taking.

The extensive body of empirical research on tournament incentives and risk taking by financial professionals is motivated, to a very large extent, by theoretical models of rank-order contests. This theoretical literature models agents who do not receive any rewards for absolute performance, clearly not even a first-order approximation of the structure of incentives for financial professionals. Thus, our work has obvious implications for the development of hypotheses for this stream of empirical research.

The most obvious implication of our analysis is that tests of the effect of tournament incentives on risk taking should, at a minimum, control for the interaction between absolute rewards and rank-based rewards. With respect to downside risk taking, our analysis predicts that the interaction effect will be negative. With respect to upside risk taking, our analysis predicts that direction of the interaction effect will depend on the relative magnitude of absolute performance versus rank rewards and the degree of ability heterogeneity between the competitors.

The empirical literature has developed measures for pecuniary rank and absolute performance rewards (e.g., Kini and Williams, 2012; Coles et al., 2018b; Ma et al., 2019). Measuring non-pecuniary rank rewards is more challenging. The theoretical literature frequently posits that non-pecuniary rewards for rank dominance are derived from the social status associated with high rankings. However, in a laboratory experiment, Kirchler et al. (2018) documents that the risk-taking behavior of financial professionals is profoundly influenced by rankings, even when rankings are anonymous, i.e., each subject is informed

of her own ranking but not the rankings of the other subjects. In contrast, student risk-taking behavior is not affected by anonymous rankings. In another laboratory experiment, Lindner et al. (2019) find that student-subject behavior is affected by publicized rankings.

This evidence suggests that some agent populations have intrinsic rank-preferences, i.e., their welfare is increased by knowing that they "are the best," even if their ranking is not communicated to other agents and is not associated with any pecuniary rewards. Other populations have rank preferences based on social status. Rank incentives based on social status depend on the degree to which agents value social dominance. Research in the social psychology literature suggests that the valuation agents place on social dominance can be predicted by the personality sub-trait aggression (Lauriola and Levin, 2001).

Under the hypothesis that the agents competing in a given contest derive similar non-pecuniary rewards from rank dominance, cross sectional variation in their behavior will be determined by differences in their pecuniary incentives. Given the discussion above, this hypothesis appears to be reasonable. Agents competing for rankings compete with agents within the same occupational group. Both the sociology literature on homophily and group formation (Ruef et al., 2003), and the social psychology literature on personality/job matching (Holland, 1973), suggest that agents within a given occupational group have similar personalities and, thus, have similar valuations of rank dominance.

Thus, it seems possible to test the predictions of the risk-taking contest model by measuring the effects of the mix between pecuniary absolute and relative performance rewards on risk taking. Developing hypotheses for such tests would require extending our analysis to consider contests where contestants receive heterogeneous rewards for rank and absolute performance. Such an extension is certainly possible, but considering this paper's length, a task for future research.

Chapter 3

Swing Pricing and Fragility in Open-End Mutual Funds

Abstract: How to avert fragility in open-end mutual funds? In recent years, markets have observed an innovation that changed the way open-end funds are priced. Alternative pricing rules (known as *swing pricing*) adjust funds' net asset values to pass on funds' trading costs to transacting shareholders. Using unique data on investor transactions in U.K. corporate bond funds, we show that swing pricing eliminates the first-mover advantage arising from the traditional pricing rule and significantly reduces redemptions during stress market conditions. The stabilizing effect is internalized particularly by institutional investors and investors with longer investment horizons. The positive impact of alternative pricing rules on fund flows reverses to some degree in calm periods when costs associated with higher tracking error dominate the pricing effect.

3.1 Introduction

Fragility in financial institutions poses a significant threat to economic stability and social welfare. Academics and policy makers have long been studying runs on banking institutions (Diamond and Dybvig, 1983); more recently, a rapid growth of shadow banking, including that of asset management companies, raised concerns that similar phenomena may also be present in the non-banking sector (Allen et al., 2009; Gennaioli et al., 2013). As experienced during the financial crisis of 2008, when market conditions unexpectedly deteriorated, investors ran on open-end funds, causing fire sales and market dislocations.[1]

While understanding the origins of fragility is certainly important, of equal importance is the question of how to mitigate it. In the banking sector, the presence of deposit

[1]Coval and Stafford (2007) study runs in equity mutual funds. Chen et al. (2010) analyze flows in bond mutual funds, while Schmidt et al. (2016) analyze runs on money market funds.

insurance and government guarantees have long been recognized as stabilizing forces. At the same time, we know much less about equally effective mechanisms in the non-banking sector, especially in the absence of explicit guarantees. Common approaches utilized by fund companies to manage redemption risk during market stress include cash buffers or redemption fees, but such tools are not as effective in practice (Chernenko and Sunderam, 2016, 2019), or can even exacerbate fragility (Zeng, 2017). In this paper, we evaluate empirically a hitherto unexplored mechanism to reduce fragility in open-end funds, *swing pricing*.

To better understand our empirical context, it is useful to outline the economic friction causing fragility, namely, the pricing mechanism used by open-end funds (Chen et al., 2010). Under the traditional pricing rule, fund investors have the right to transact their shares at the daily-close net asset value (NAV) of the fund portfolio. As a result, the price that a transacting shareholder receives does not take into account the corresponding transaction costs that may arise because portfolio adjustments associated with shareholder transactions typically take place over multiple business days following the redemption request. Thus, the costs of providing liquidity to transacting shareholders are borne by non-transacting investors in the fund, which dilutes the value of their shares. Chen et al. (2010) show that this mechanism can produce a first-mover advantage and amplify the impact of negative shocks, especially during market-wide stress when market liquidity drops and in environments in which strategic complementarities are important.

Alternative pricing rules—typically known as *swing pricing* or *dual pricing*—aim to adjust funds' net asset values so as to pass on the costs stemming from transactions to the shareholders associated with that activity. In this paper, we conduct a systematic empirical analysis to evaluate the impact of swing pricing on the dynamics of fund flows. We ask: To what extent does swing pricing help funds to retain investor capital during periods of market stress? Are funds able to prevent dilution in fund performance and eliminate first-mover advantage? How do individual fund investors respond to fund companies' swing pricing rules?

Alternative pricing rules take three different forms. The first one is *full swing pricing*, whereby a fund's net asset value (NAV) can be adjusted up or down on every trading day in the direction of net fund flows: If net flows are positive the NAV shifts up and if net flows are negative the NAV shifts down. The magnitude of the shift is known as the *adjustment factor*. The second form, the *partial swing pricing*, is invoked only when net flows cross a pre-determined threshold, namely the swing threshold. For both forms, a single price applies to all transactions including both redemptions and subscriptions. The third form, referred to as *dual pricing*, is similar to full swing pricing in that a fund's NAV can be adjusted on every trading day without a requirement to cross the threshold. However, it differs in that a fund trades at two prices—subscribing investors purchase their

shares at the NAV adjusted up (ask price) and redeeming investors redeem their shares at the NAV adjusted down (bid price).

Funds are permitted, but not required, to use the alternative pricing structures and they have a full discretion for the values of adjustment factors. Investors only know if the fund applies alternative pricing rules, but they do not know the precise values of adjustments. They learn about them from the *ex-post* transaction prices. In this regard, the observed flow dynamics results from an interplay between managers' ability to assess illiquidity costs in the market and investors' learning about managerial pricing decisions.

Regulation permitting swing pricing rules has become effective in the U.S. only in November 2018; however, these rules have been used in several European jurisdictions over the past few decades. To analyze the impact of the rules, we obtain data on corporate bond open-end funds that fall under the supervisory jurisdiction of the Financial Conduct Authority (FCA). Choosing bond funds as a testing ground allows us to capture a significant component of a fund portfolio illiquidity, a key determinant of fund fragility. The data have a number of unique features. Most relevant, we obtain detailed information on funds' pricing practices, including the daily adjustment factor. Moreover, we observe the holdings of the funds' end-investors, which allows us to look at individual-specific responses to pricing rules and address the identification concerns. Finally, the data cover a period from January 2006 to December 2016, which includes a number of high-stress episodes, such as the 2008 global financial crisis, the European debt crisis, or the Taper Tantrum. Periods with market-wide stress are natural candidates to study the risk of fund fragility. In our study, we measure market stress using abnormal values of option-implied volatility index (VIX).

We begin our empirical analysis by examining the determinants of the dilution adjustment factor. If the pricing rules matter, we should expect fund companies to implement adjustments in times of high market stress when aggregate liquidity tends to be low. This is precisely what we find. The adjustment factor is substantially higher (for instance, it nearly quadruples during 2008 crisis) in periods of higher portfolio illiquidity and periods of market stress.

We next investigate whether swing pricing affects the level of fund flows during market stress. While academic research on swing pricing is scarce, there is an ongoing debate about it among practitioners and supervisory authorities. One view is that swing pricing can mitigate fragility in funds by removing the negative externalities arising from investor flows.[2] An alternative view suggests that swing pricing can increase fragility. Anticipating

[2]Blackrock Viewpoint Series titled *Fund structures as systemic risk mitigants* (2014).

an increase in near-term future liquidation costs, investors may become more sensitive to negative shocks.[3]

We find that funds with traditional pricing rules experience significant outflows during market stress, in line with prior literature (Mitchell et al., 2007; Ben-David et al., 2012). Importantly, this effect *almost completely* reverses for funds that adopt swing pricing, lending support to the view that such rules reduce fragility risks. All results are robust to including a range of fixed effects (e.g., fund family, investment style, region of sale), front and back-end loads, and alternative definitions of market stress. We also examine each type (full swing, partial swing, and dual pricing) separately and observe largely similar patterns.

The economic magnitude of the average effect is fairly sizable. During stress periods, an average traditional fund loses capital worth of £8.86 million every month; the corresponding loss for funds with alternative pricing is only £0.2 million. Moreover, evidence from quantile regressions shows that the magnitude of the effect substantially increases for a subset of funds with the highest flow sensitivity, supporting prior evidence on the heterogeneity of funds' fragility (Schmidt et al., 2016).

A potential concern with the interpretation of pricing rules causing flows is that funds and investors with different characteristics may self-select into different pricing structures. A significant advantage of our data is their unique granularity that allows us to tackle this concern, and thus to pin down the economic mechanism behind our findings. These elements of the data make the paper uniquely suited to examine the impact of systematic shocks at the individual investors' level, different from the previous studies of asset management firms that exploit share-class-level data, thus assuming homogeneity within a particular group of investors (e.g., Kacperczyk and Schnabl, 2013; Schmidt et al., 2016).[4]

To this end, we first identify a subsample of funds which switch their pricing methods from traditional to alternative within our sample period, and we examine individual (same) investors' behavior before and after the switch. Empirically, we match the sample of switchers to non-switchers along various characteristics and estimate the treatment effect at the end-investor level. We employ a triple-difference test in which we compare investor flows of switchers vs. non-switchers before and after the switching conditional on the level of stress in the aggregate market. We additionally take advantage of end-investor fixed effects, which allows us to study the behavior of the same investor before and after the

[3]Cipriani et al. (2014) provide a theoretical model of preemptive runs when intermediaries impose gates or redemption fees.

[4]To our knowledge, the only other paper that studies runs with that level of data granularity is Iyer and Puri (2012). However, their objective is to trace a banking panic in one specific Indian bank, whereas we focus on the question of how to mitigate the threat of runs in the asset management sector.

change. The staggered nature of switching dates is helpful in capturing the effects due to switch.

Our findings provide strong evidence that results are not solely due to selection; pricing structures also alter investor behavior. We find that, the same investor is significantly less likely to redeem her shares in a stress period when a fund uses swing pricing than when the fund uses traditional pricing. Moreover, funds that switch to alternative pricing structures attract less flows outside stress periods thereafter. For a limited sample of investors, we also address a possible endogeneity concern due to time-varying investor-specific omitted variables by showing that the differential effect across the two structures is similar when we compare the behavior of the *same* investor in two different funds, one of which switches the structure.

We complement the baseline analysis with a series of tests that provide insights into the economic mechanism behind the results. First, we examine flow–performance sensitivity and show that swing pricing does not have a significant impact on the sensitivity of investor inflows to good performance, but it significantly reduces the sensitivity of outflows to bad performance. The asymmetric nature of the response strongly supports the interpretation that swing pricing mitigates fragility arising from costly asset liquidations. Moreover, consistent with these funds being more resilient to stress events, we find that they are less likely to be terminated compared with funds with the traditional pricing rule.

Next, we zoom in on cross-sectional differences. Consistent with the predictions of Chen et al. (2010), the mitigating role of alternative pricing rules is particularly important for funds with illiquid assets and dispersed ownership, that is, funds that are most fragile. When we examine the differences across investors, first, we find that it is the institutional investors who alter their behavior the most after the fund switches to swing pricing. Institutional investors sell heavily when the fund use the traditional pricing rule (thus retail investors bear the costs due to their transactions), and after the switch, they significantly reduce their redemptions. This is consistent with the idea that, being more sophisticated, these investors better understand the negative externalities arising from funds' pricing practices. In a similar vein, investors with longer investment horizons also alter their behavior and sell less after the switch, as swing pricing is a tool to protect the value of their shares. These findings strongly support that strategic complementarities, rather than mechanical rebalancing rules of unsophisticated investors, drive our findings.

We further evaluate the consequences of funds' pricing methods. The negative consequence of the traditional pricing rule is the dilution effect due to large outflows for non-transacting investors, which gives rise to the first-mover advantage. If our findings are driven by the pricing structure, we expect alternative funds to be able to remove the first-mover advantage arising from fund outflows. We find that outflows indeed negatively impact

subsequent fund performance for funds using the traditional pricing rule. However, the negative impact of outflows on fund performance almost completely dissipates for funds with alternative pricing. This finding provides evidence that funds use the alternative pricing structures effectively enough to eliminate the first-mover advantage arising from the traditional pricing rule.

While swing pricing is beneficial for funds in that it reduces redemptions during market stress, these funds have smaller inflows in other periods. One reason is that dilution adjustment in fund prices can increase a fund's tracking error as fund prices are adjusted to pass on the trading costs to transacting shareholders. Attracting new investors could then become more difficult. We find that funds with swing pricing indeed have higher tracking errors and investors strongly respond to funds' tracking errors in their investment decisions. Consequently, such funds attract fewer new investors, on average.

In the final set of results, we test whether funds with swing pricing rules tend to treat this tool as a substitute to other means of liquidity risk management, such as cash holdings, portfolio diversification, or fund loads. We find that such funds hold less cash compared to funds with traditional pricing. The effect for portfolio diversification and fund loads is statistically insignificant. One reason why load fees may not be as effective is that they do not eliminate the first-mover advantage as proceeds from loads are not retained in the fund; instead, they are used to compensate brokers for their services (Chen et al., 2010).

Related literature

From a broad perspective, our paper contributes to a vast literature on stability of and runs on financial institutions. The main focus of this literature has been mostly the banking sector. Recent body of work acknowledges that non-banking institutions, such as mutual funds, can also destabilize the markets. In particular, several papers document significant declines in open-end fund performance due to *aggregate* fund outflows and suggest that the resulting dilution in fund performance can lead to fragility (e.g., Edelen, 1999; Coval and Stafford, 2007; Alexander et al., 2007; Christoffersen et al., 2006). Our focus instead is to show a mechanism that mitigates fragility using *disaggregated*, investor-level data.

Chen et al. (2010) build a global game model and show that the traditional pricing rule used by open-end funds can lead to runs on funds because predictable declines in NAV following fund outflows generate a first-mover advantage. Consistent with the predictions of the model, they document that the flow-to-performance relationship is stronger for funds investing in less liquid stocks. Goldstein et al. (2017) echo the message by showing that

corporate bond funds exhibit a concave flow-to-performance relationship. Our paper supports this mechanism by showing the importance of first-mover advantage and illiquidity in the corporate bond fund sector through the lens of swing pricing.

A related literature discusses possible remedies to runs in open-end funds with cash being the most natural candidate. Morris et al. (2017) explore the cash hoarding channel and argue that funds sell more assets than required to cover outflows. Chernenko and Sunderam (2016, 2019) analyze the cash-cushioning approach and conclude that funds' cash holdings are not sufficiently large to eliminate fire sales. One theoretical explanation behind this finding is Zeng (2017) who argues that cash management cannot prevent runs; instead, cash usage can actually exacerbate the runs on open-end funds.

We offer an alternative tool to mitigate run risks that gets at the core of the friction, the pricing mechanism. Swing pricing, which allows for dilution adjustment on fund NAV, reduces the first-mover advantage arising from the traditional pricing and substantially reduces the outflows during crisis periods. In this respect, our findings are consistent with the recent theoretical study of Capponi et al. (2020) who show the stabilizing effects of swing pricing. Our paper corroborates their predictions empirically and provides additional cross-sectional and time-series tests of the theory.[5]

3.2 Institutional Background

Open-end funds provide daily liquidity to their shareholders. Typically, on any given day, fund investors have the right to transact their shares at the daily-close NAV. However, trading activity and other changes in portfolio holdings associated with shareholders' transactions may occur over multiple business days following the transaction requests; hence, the costs of providing liquidity to transacting shareholders can be borne by non-transacting fund investors. Such costs reduce fund performance, thereby diluting interests of non-transacting shareholders.

To address the dilution effect arising from transacting shareholders' flows, alternative pricing rules have emerged which allow open-end mutual funds to adjust their NAVs. These rules, known as *swing* or *dual* pricing, exist in many European domiciles: Luxembourg, Finland, France, Ireland, Jersey, Norway, Switzerland, and the U.K. All registered open-end investment companies in the jurisdictions have been eligible for such pricing over the

[5]Lewrick and Schanz (2017a) compare funds domiciled in Luxembourg, where funds are permitted to use swing pricing, to U.S.-domiciled funds during the period when they were not allowed to use swing pricing. Their data span a short period, which does not include a major stress period. Importantly, Lewrick and Schanz (2017a) do not observe the cross-sectional variation in funds' pricing rules—they simply compare all Luxembourg funds (including both swing and traditional funds) to the U.S. funds (traditional funds). Perhaps not surprisingly, authors do not find important differences.

past few decades. In the U.S., the Securities and Exchange Commission (SEC) adopted rules permitting funds to use the new pricing in 2016. They have become effective in November 2018.[6]

Two main alternative pricing mechanisms are employed in European jurisdictions: swing pricing and dual pricing. When a fund uses swing pricing, NAV is moved up or down, depending on whether the fund faces a net inflow or a net outflow: NAV swings up if a fund gets a net inflow, and swings down in case of a net outflow. The size of the swing, known as a *swing* or *adjustment factor*, while at the discretion of fund managers, should compensate non-transacting shareholders for the costs of trading due to capital activity by transacting shareholders.[7] Fund managers typically use either of the two types of swing pricing: partial swing pricing or full swing pricing. Partial swing funds move the price only when the net fund flow is greater than a pre-determined threshold, the *swing threshold*. This threshold is usually set in terms of a percentage or basis point impact, and to avoid any potential gaming behavior by investors, it is not publicly disclosed. Full swing funds can swing their prices every day. The direction of the swing can depend on the direction of the daily fund flow or it can be set on a long-term basis based on expected flows.[8] In both types of swing pricing, the final price applies to all transacting shareholders (whether they are redeeming or subscribing).

Different from swing funds, which trade at a single price, dual priced funds trade at two separate prices, bid and ask. Investors purchase them at the ask price and sell at the bid price. Depending on the net fund flows, a fund manager can adjust the spread between a fund's bid and ask prices up to the bid-ask spread of the fund's underlying assets.[9] Proceeds from net inflows or net outflows are reinvested in the fund, which protects non-transacting shareholders from dilution.[10] Compared with swing funds that do not disclose their adjustment factor, dual priced funds are more transparent as both bid and ask are publicly available.

Funds are permitted, but not required, to use dilution adjustments. Although no explicit regulation stipulates to do so, several swing funds choose to cap their swing factors (often self-impose a cap of 2%). The pricing rule is typically determined at the start of the fund,

[6]Other countries allowing swing/dual pricing are Australia, Cayman Islands, and Hong Kong.

[7]Rules permit funds to swing NAV only to address the dilution effect arising from investor flows. To address stale prices, managers can use evaluated prices provided by third-party data providers.

[8]For full swing funds, direction of daily swing factors lines up with the direction of daily flows 85% of the time.

[9]The final price can include sales charges, if any. Sales charges are not common, and importantly, they are not retained in the fund. We calculate the spread in dual funds' bid and ask prices before any sales charges.

[10]Recently, FCA recognized that managers of some dual-priced funds were retaining the profits from the spread on days when inflows and outflows net out (so called 'box profits'). The new rules, which became effective on April 1, 2019, require fund managers to return box profits to the fund investors. See `https://www.fca.org.uk/publication/policy/ps18-08.pdf`.

and the dilution adjustment is applied uniformly across all shares. If a fund uses swing or dual pricing, it must disclose this information in its prospectus; however, funds are not required to report swing factors and swing threshold. Investors only observe the final price.

Funds are required to ensure an equitable treatment of their investors. To oversee the use of dual/swing pricing, most funds set up valuation and pricing committees, either as a standalone committee or as part of the funds' boards. Moreover, depositary banks, which in the E.U. provide fiduciary and custodian services to investment funds authorized to be marketed in any E.U. jurisdiction, oversee the affairs of the funds, including those related to pricing. Depositary banks are obliged to ensure that the fund complies with the rules and its own constitutional documents. Most depository banks in the E.U. are custodian banks such as Barclays, JP Morgan, Goldman Sachs, HSBC, and State Street Corporation. Depository banks are prohibited from overseeing funds that belong to the same financial institution—that is, for instance, Goldman Sachs is not allowed to oversee the mutual funds offered by Goldman Sachs. However, it is possible for depositary banks to oversee multiple investment funds from the same financial group.

Alternative liquidity management tools are, in principle, available to fund managers; however, these alternative tools are not commonly used in practice. For instance, funds can introduce redemption gates (deferring redemptions to a future valuation point) or use redemptions in kind (returning a slice of the portfolio instead of returning cash to redeeming shareholders). Such measures are rarely used. Importantly, the potential availability of alternative tools go against finding significant effects due to pricing rules. In addition, funds can aim to manage liquidity risk by maintaining buffers of cash and cash equivalents, such as Treasury bills and commercial papers. Holding cash and cash equivalents, however, can be associated with important opportunity costs. Moreover, a recent study by (Zeng, 2017) casts doubt on the effectiveness of cash and cash equivalents in mitigating fragility in funds. Whether alternative tools are substitute to alternative pricing is an empirical question that we examine in Section 3.4.8.

3.3 Data

3.3.1 Sample construction and measures

We obtain our data through a request sent by the FCA to major UK based asset management companies with corporate bond fund offerings.[11] The FCA requested data on

[11] This also includes U.K. subsidiaries of non-U.K. asset management companies.

all corporate bond mutual funds domiciled in the U.K. or whose investment management decisions are taken from the U.K.[12] Through this data request, the FCA received data on 299 corporate bond mutual funds (including dead funds) from 24 asset management companies.[13] A fund is defined to be a corporate bond fund if at least 50% of its portfolio is invested in corporate bonds; however, the majority of funds in our sample have bond holdings of more than 80%. The data include funds from leading U.S. and European multinational asset management companies, covering the period from January 2006 to December 2016.

The FCA database has several unique features. First, it includes comprehensive information on funds' dilution adjustment practices. We observe fund NAVs, prices, swing factors, and swing thresholds at daily frequency. While funds are required to disclose the type of pricing rules that they use, they are not required to disclose swing factors and thresholds to the public. For dual funds, we also observe the daily bid and ask prices. An additional unique feature of our data is information on end-investors' holdings (at monthly frequency) and their investment type (retail vs. institution). We also observe various fund-level characteristics, such as total net assets (TNA), returns, cash, and asset holdings. We complement the FCA data with information from Morningstar on fund fees (expenses) and institutional class indicators.

Since pricing rules are applied uniformly across all share classes, we follow the literature (e.g., Kacperczyk et al., 2005) and aggregate observations to the fund level. For qualitative attributes (year of origination and country of domicile), we use the observation of the oldest class. For fund size (total assets under management), we sum the TNAs of all share classes. We take the TNA-weighted average for the rest of the quantitative attributes (e.g., returns, alphas, and expenses).

Through the matching of various databases, we arrive at the final sample that includes 224 open-end corporate bond mutual funds in 22 families that are open to new and existing investors. The sample excludes ETFs, money funds, and index funds. As of December 2016, 18% of the funds apply traditional pricing, and the remaining 82% use alternative pricing. Within the latter group, 18% and 54% use full and partial swing pricing, respectively. The remaining 10% use dual pricing. Over the period 2006-2016, 34 funds switched their pricing schemes from the traditional to alternative structures, specifically to swing pricing. We do not observe any switches from alternative to traditional pricing scheme

[12]The latter condition selects funds that have a significant presence (usually office) in the U.K. Funds in our sample are domiciled in various jurisdictions, the majority of which are in the United Kingdom, Luxembourg, and Ireland, representing, 55%, 31%, and 11% of the sample, respectively.

[13]20 funds offered by four asset management companies with combined assets under management of about £3.4bn (as of the end of 2016) failed to respond to the data request, a relatively small portion of the overall sample.

during our sample period, which suggests that the market may favor alternative pricing but is likely still in a transition phase.

We conduct our baseline analysis at monthly frequency. For each fund-month observation, we define a number of variables. *Flow* is the monthly change in the quantity of shares outstanding multiplied by the share price, divided by a fund's TNA. Both the numerator and the denominator are measured as of time t to prevent a potential contamination in *Flow* due to fund price adjustment. Notably, our measure is based on directly observed flows rather than on indirect measures imputed from fund size as is common in the literature.[14] *Return* is the fund's monthly raw return net of expenses. Following the earlier studies on corporate bond mutual funds (e.g., Goldstein et al., 2017; Choi and Shin, 2018), we estimate fund *Alpha* using a 12-month rolling-window regression model of monthly excess returns on excess aggregate bond market and aggregate stock market returns. We obtain market indexes from Barclays. *Size* is the natural logarithm of a fund's TNA; *Age* is the natural logarithm of a fund's age, in years; *Expense* is a fund total expense ratio; *Inst* is the fraction of a fund's assets held by institutional investors. *Illiquidity* is the value-weighted average of bid-ask spreads of a fund's assets.[15] Bid-ask prices are obtained from Thomson Reuters Datastream.[16] We winsorize all variables at the 1% level. We provide details on variable definitions in Appendix Chapter A.

We follow the literature and define *Stress* as an indicator variable equal to one if the average of the end-of-day Chicago Board Options Exchange Volatility Index (VIX) is above the 75^{th} percentile of the sample in a given month. Within our sample, *Stress* covers the episodes of 2008 global financial crisis, the European debt crisis, the downgrade of the credit ratings of U.S. federal government, and the Taper Tantrum. Figure C.1 shows the time series of VIX during our sample period. Later in the paper, we show the robustness of our results to alternative definitions of *Stress*.

3.3.2 Descriptive statistics

Table C.1 presents the descriptive statistics for the fund characteristics in our sample. For brevity, we categorize funds into two groups: funds that use the traditional pricing

[14] Our results are robust to using the traditional flow measure in which the denominator (fund size) would be measured in $t-1$, and the numerator would be inferred from changes in fund size from $t-1$ to t.

[15] Since we do not have data on intraday bond returns, we use daily bid-ask spreads. Our main variable of interest is fund flows. We mostly use *Illiquidity* as a control variable.

[16] When available, we use Thomson Reuters' composite price, which is an average price from multiple pricing sources. When composite price is missing, we use the evaluated price, which is provided daily by the Fixed Income Pricing Service team at the Thomson Reuters. This pricing service uses proprietary evaluation models and is used by many industry participants, e.g. for NAV calculations. If this price is also missing, we use prices provided by iBOXX or ICMA.

rule versus those with alternative pricing rules (swing or dual). Panel A and B shows the descriptive statistics for funds with alternative and traditional pricing rules, respectively.

Table C.1 shows that funds with traditional pricing appear to be similar to those with alternative pricing in a number of ways. First, they have similar TNAs. The average size for funds with alternative pricing is £141 million while the corresponding number for funds with traditional pricing is £143 million. Further, the two groups have similar expenses, with an average annual expense ratio of 0.88% for funds that use the traditional pricing and an average expense ratio of 0.75% for funds with alternative pricing. Funds with alternative pricing appear to be slightly older (7.92 vs. 5.75 years). In general, along many characteristics, our sample is quite similar to the U.S. corporate bond funds analyzed by Goldstein et al. (2017).

In the last two columns, we report the descriptive statistics on asset illiquidity and investor type for the two groups of funds. Funds with alternative pricing hold more illiquid assets. On average, the value-weighted bid-ask spread of the funds' assets is about 94 bps while it is 80 bps for funds with traditional pricing. This finding is consistent with the hypothesis that funds with more illiquid assets are more fragile and thus are more likely to use alternative pricing to offset it. Further, ownership by retail investors in funds with alternative pricing tends to be higher (77% vs. 66%). The ownership structure is important because investors with different levels of sophistication are likely to internalize the negative externalities arising from traditional structure differently. We examine this in Section 3.4.5.

3.4 Empirical Results

3.4.1 Dilution adjustment factor across funds and time

We start our analysis by examining the time-series patterns in dilution adjustment factors. Alternative funds are permitted to adjust their NAVs to account for trading costs arising from price impact, bid-ask spreads, and other explicit trading costs (e.g., stamp duty, taxes). We define *Adjustment Factor* as the daily *absolute value* of swing factor for swing funds. For dual funds, it is equal to the half spread of the funds' bid and ask prices, $0.5 * (ask - bid)/mid$. During our sample period, *Adjustment Factor* of funds with full swing and dual pricing is approximately 33 bps. For partial swing funds, the median *Adjustment Factor* is zero because swinging is invoked only when daily net flows cross a specific threshold. As reported in Table C.14 of the Internet Appendix, the most commonly used thresholds (in absolute terms) are 1% and 3%.[17] 90% of partial swing funds use

[17] These thresholds approximately correspond to 5% and 10% tails of the daily net flow distribution.

thresholds that are less than 3%. The average dilution adjustment factor for partial funds is 57 bps once we restrict our sample to days with non-zero factor values.

Figure C.2 shows the time-series variation in average *Adjustment Factor* of swing and dual pricing funds. The adjustment factor is relatively small outside the crisis periods, varying from 18 bps to 25 bps, but it substantially increases in adverse market conditions. For example, the average factor spikes up—nearly quadruples—during the 2008 global financial crisis; similarly, adjustment factors are at relatively high levels during the European debt crisis. Overall, patterns in the average factor line up with those documented in other studies. Among others, Biais and Declerck (2007) document that, outside the crisis periods (from 2003 to 2005), effective spreads in European corporate bonds ranged between 12 bps and 22 bps. Also, Dick-Nielsen et al. (2012) document dramatic increases in corporate bond illiquidity measures (such as price impact and bid-ask spreads) during 2008.[18]

Next, we analyze which fund characteristics are associated with dilution adjustment factor. Since we do not observe detailed order and transaction data, estimating funds' trading costs is difficult. Because trading illiquid assets is more costly than trading liquid assets, we expect the degree of illiquidity of a fund's assets to be an important determinant of its adjustment factor. Moreover, because trading costs tend to surge during stress market conditions, we expect the adjustment factors to be particularly high during such periods. To test these predictions, we estimate the following regression model:

$$AdjustmentFactor_{i,d} = \alpha + \beta_0\, Illiquidity_{i,d} + Day\ FE + Fund\ FE + Other\ Fund\ Characteristics_{i,d} + \varepsilon_{i,d}, \quad (3.4.1)$$

where *Illiquidity* is the daily value-weighted average of the bid-ask spread of fund i's assets, *Day (Fund) FE* are day (fund) fixed effects. To assess the role of other fund characteristics, we extend the model to include *Size*, *Age*, *Expense*, and *Inst*, all measured at the end of the previous month. Furthermore, in the latter specifications, we remove day fixed effects and include *Stress* (VIX is above the 75$^{\text{th}}$ percentile of the sample) to capture the time-series variation. We cluster standard errors by fund and day.

We report the results in Table C.2. In column (1), we present results from estimating the univariate regression model with *Illiquidity* as the main explanatory variable. In columns (2)-(3), we sequentially add other fund characteristics and fund fixed effects. Across all specifications, we find that *Illiquidity* is significantly positive, indicating that asset illiquidity is an important determinant of funds' adjustment factors. Besides *Illiquidity*, other

[18]To assess trading costs, funds typically use a measure known as implementation shortfall, which is analogous to effective spread. Other costs, such as commission fees are often waived, and stamp duty and taxes make up about 5 bps (e.g., Busse et al., 2017). Anecdotal evidence suggests that funds often obtain trading cost estimates from third-party data providers such as Thomson Reuters, Markit, and others.

fund characteristics do not appear to have an important explanatory power. In columns (4) and (5), we show the results with *Stress* as the main explanatory variable. Consistent with patterns observed in Figure C.2, the adjustment factor significantly increases during periods of market stress. Finally, in column (6), we present the results from the model in which we interact *Illiquidity* and *Stress*. The results indicate that the adjustment factor is particularly high for illiquid portfolios during stress periods, as one would expect.

3.4.2 Fund flows and alternative pricing: Cross-sectional evidence

Under the traditional pricing, fund investors have the right to redeem their shares at the fund's daily-close NAV. Following substantial outflows, a fund needs to adjust its portfolio and consequently it may conduct costly and unprofitable trades. Since most of the resulting trades are likely to be executed *after* the day of redemptions, such costs are not reflected in the NAV paid out to redeeming investors but are rather borne by those who stay in the fund. Chen et al. (2010) show that this mechanism can produce a first-mover advantage and amplify the impact of negative shocks, especially during market-wide stress when market liquidity dries up. To the extent that alternative pricing protects the interests of remaining investors by passing on the trading costs to redeeming investors, such fragility can be mitigated.

Formally, we evaluate the impact of alternative pricing on fund flows by estimating the following regression model:

$$Flow_{i,t} = \alpha + \beta_0\, Alternative_{i,t} + \beta_1\, Stress_t + \beta_2\, Alternative_{i,t} \times Stress_t + \beta_3\, Controls_{i,t} + \varepsilon_{i,d}, \quad (3.4.2)$$

where *Alternative* is an indicator variable which equals one if a fund is using one of the alternative pricing mechanisms. *Flow* and *Stress* are defined as before. Control variables include lagged fund characteristics (measured previous month-end) such as *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. We cluster standard errors by fund and month.

In Table C.3, Panel A, we report the results of OLS regression. In column (1), we report the results for the univariate regression and in column (2) we report the results for the regression model with fund controls. In both specifications, the coefficient of *Alternative*×*Stress* is positive and statistically significant. Moreover, the value of the coefficient nearly cancels out the negative value of the coefficient of *Stress*. For instance, in column (1), the coefficient is 1.04 and that of *Stress* is -0.99. These results indicate that alternative pricing is effective in reducing outflows in bad times. At the same time, we also find that

the coefficient of *Alternative* is negative, though statistically insignificant, which suggests that alternative funds have less inflows than traditional funds in good times.

To the extent that funds with different pricing rules may have different characteristics, our test sample in columns (1)-(2) may be unbalanced. To sharpen the interpretation of our findings, we match each of our swing funds to the sample of funds which rely on traditional pricing. Following Loughran and Ritter (1997), we find the nearest bond fund using a matching algorithm which minimizes the sum of the absolute percentage differences in lagged values of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. We perform the matching with replacement. If a fund is selected as a suitable match to more than one fund, we use this observation only once.

In columns (3)-(7), we present the results based on the matched sample. In column (3), we repeat the same estimation as in column (2). In column (4), we include fund fixed effects to account for time-invariant omitted fund characteristics. In column (5), we include time fixed effects. In column (6), we include family fixed effects; in column (7), style fixed effects. The findings reported across the specifications appear robust. Results are both statistically and economically more significant when we use the matched sample.

In Table C.2 of the Internet Appendix, we provide extended robustness tests using additional fixed effects (such as fund's location domicile, region of sale, investment area), front and back-end loads, and alternative definitions of market stress, which are defined based on TED spread, LIBOR rate, and Merrill Lynch's MOVE index. Results are similar throughout.

Estimates imply that, during stress, traditional funds lose, on average, capital worth of £8.86 million in each month. The corresponding loss for funds with alternative pricing is only £0.2 million. For the matched sample, the difference is even larger (£10.32 million vs. £0.1 million). Given that the average fund has £150 million in assets under management, the average effects may not seem large. We note, however, a significant variation in outflows in the cross-section of funds. To show this effect directly, we resort to estimating quantile regression models. We present the results in Panel B of Table C.3.

Results obtained from the OLS regression are similar across different percentiles ranging from the 50^{th} to the 0.5^{th}, that is, the negative impact of *Stress* on fund flows is almost fully reversed for alternative funds. Importantly, the economic magnitudes increase substantially as we move from the median fund towards the funds in the left tail of the flow distribution. For example, the estimates increase *five* times when we compare the mean response to the response of a fund in the 5^{th} percentile. Overall, results show that the economic significance of alternative pricing rules is particularly large in the tails of the flow distribution.

In Table C.4, we decompose the effect of the alternative pricing into specific sub-components (full swing, partial swing, and dual pricing). This decomposition allows us to understand better the information content of alternative pricing rules. While dual pricing is generally known to investors, the magnitude of the swing factor is unobserved. Given that we observe funds with different structures, we can empirically assess how important the unobservability is for investors in such funds. The results indicate the issue does not seem as relevant. We find similar patterns (that the negative impact of stress market conditions on fund flows almost completely dissipate) for each individual component. In statistical terms, results are strongest for partial swing and weakest for dual priced funds, arguably reflecting the differences in statistical power in each subsample. As noted in Section 3.3, majority of alternative funds use partial swing while the dual pricing form is rare.

3.4.3 Fund flows and alternative pricing: Evidence from switching funds

One potential concern with interpreting the results in Section 3.4.2 is that cross-sectional differences in flows to funds with different characteristics may reflect underlying differences across funds with different structures or results may reflect self-selection of funds into different structures. While including fund controls alleviates this issue, it is unlikely to solve it fully.

In this section, we address this issue by taking advantage of a subsample of funds which change its pricing method. Over the period 2006-2016, 34 funds from six asset management companies switched their pricing schemes from the traditional to alternative structures.[19] Panel A of Table C.16 lists the dates when the switch took place.

Anecdotal evidence from interviews with the companies suggests that the switches were unlikely to be related to fund performance, flows, or other characteristics correlated with flows. Since some of the funds within the same families did not change their structures, it is also unlikely that the switches were purely family-wide decisions. Finally, the staggered nature of the switches makes it unlikely that the change in structure reflected a structural aggregate change in the market.

Evidence from funds' responses

In order to assess the impact of the change in fund structure on fund flows, we first look at the results at the fund level. Our empirical strategy involves comparing the flows

[19] We do not observe any switches from alternative to traditional pricing scheme during our sample period.

of funds that change their structures to alternative pricing—treated funds—before and after the change. For our analysis, we specify a window of 48 months, with 24 months before and 24 months after the reported switch date. Because the observed effect in flows could be correlated with an unobserved time effect, ideally, we would like to observe the counterfactual fund behavior in the absence of treatment. Obviously, such counterfactual cannot be observed in the data. We instead approximate the counterfactual with a control fund defined as a close match using the algorithm in Section 3.4.2.

We examine the potential pre-trends that may determine any differential response to the shock of our interest. The presence of such pre-trends cannot be tested directly; however, we can inspect their plausibility using graphical presentation and regression evidence. In Figure C.3, we present the time-series dynamics of average values for various fund characteristics around the event time. We do not observe important differences. In Panel B of Table C.16, we formally examine this by testing the differences in fund characteristics for treated funds relative to control funds before and after the event. We estimate a difference-in-differences regression model of the following form:

$$Characteristic_{i,t} = \alpha + \beta_0 Treated_i \times Post_t + \beta_1 Post_t + \beta_2 Treated_i + \varepsilon_{i,d}. \quad (3.4.3)$$

We define an indicator variable *Post* that equals one for the period after the change and equals zero before the change. *Treated* is an indicator variable that equals one for all funds that have changed their structure, and zero for the funds in the control group. Columns (1) to (7) of Panel B in Table C.16 show the results for *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, *Inst*, and *#Inv*, respectively. Across all characteristics, we find that both β_2 and β_0 coefficients are statistically insignificant, which supports no significant differential pre-trends between the treatment and the control group during the event window. The switch itself, on average, does not induce significant differential responses in fund characteristics that could predict any heterogeneous effects for fund flows after the switch.

We next evaluate the impact of change in fund's pricing structure on fund flows conditional on the level of stress in the market, similar to our specification in (3.4.2). Specifically, we estimate:

$$\begin{aligned} Flow_{i,t} = \alpha + \beta_0 Stress_t \times Post_t \times Treated_i + \beta_1 Stress_t \times Post_t + \\ \beta_2 Stress_t \times Treated_i + \beta_3 Treated_i \times Post_t + \beta_4 Post_t + \\ \beta_5 Treated_i + \beta_6 Stress_t + \beta_7 Controls_{i,t} + \varepsilon_{i,d}. \end{aligned} \quad (3.4.4)$$

Our coefficient of interest is β_0. We present the results in Table C.5. In column (1), we report the results for the specification that does not include any controls or fixed

effects. We find a strong positive and statistically significant differential effect on treated funds during market stress. The coefficient, β_0, is positive and statistically significant. Moreover, in the absence of stress, the difference in flows between treated and control group, measured by the coefficient β_3, is negative. In column (2), we add the same control variables as in Section 3.4.2. The coefficient β_0 remains positive and statistically significant. In column (3), we further include fund fixed effects to account for any time-invariant fund characteristics, while in column (4) we include time fixed effects. In both cases, the coefficient β_3 is positive and statistically significant.

Evidence from investors' responses

Our fund-level analysis based on switchers helps to trace down the effect of pricing schemes on fund flows; however, one remaining identification concern is related to investor heterogeneity. Thus far, we assume that the investor base in the treated funds remains unchanged following the treatment and any estimated differences reflect the change due to pricing rule only. However, it is quite possible that the shock itself also induces a change in the composition of investors in treated funds, and funds before and after the treatment are owned by investors with different preferences for risk or investment horizon. This concern generally applies to all large-sample studies of delegated asset management and has been difficult to address due to data limitations.[20] In this study, we are uniquely positioned to address this issue because we can observe investment decisions at the individual investor level. Consequently, we can track the behavior of a *given* investor both before and after the change in a fund's pricing rule, conditional on the overall market stress.

We begin by presenting the overall patterns in the end-investor data.[21] The average (median) fund in our sample has 596 (85) investors, 13% of which are institutional clients. An individual investor's participation in a given fund (defined as the investor's value of holding divided by total fund size) is typically small. For retail clients, the mean and median values are 0.005 and 0.003, respectively. For institutional clients, these values are larger, with the mean equal to 0.17 and median is 0.006. The maximum investor ownership in the sample is 2.3%.

[20] To our knowledge, the best treatment of this issue to date was to study differences in flows of funds with the same underlying fund portfolio but different share classes catering to various investor types (e.g., Kacperczyk and Schnabl, 2013; Schmidt et al., 2016). However, these studies make an implicit assumption that each share class has either homogeneous or stable pool of investors, which need not be true in the data.

[21] Detailed investor-level data are available for 230 funds in 20 families (vs. 299 in the full data); this number is reduced to 196 (vs. 224) if we constrain our sample to observations with full record of all variables used in our tests. The average fund excluded from investor-level analysis does not appear significantly different from the average fund included in the analysis. The client classification is provided by funds. Potential measurement error in the classification would likely go against finding significant differences between different groups.

In total, we observe about 120K investor trades taking place in our event window. We calculate the frequency of trading by each investor during the event period. We first define an indicator variable *trade* which equals one if the investor is trading in a given month. We then define *trading frequency*, which is the average of *trade* for a given investor during the event period. Cross-sectional average (median) of *trading frequency* is 0.33 (0.16).

The 25^{th} and the 75^{th} percentiles for *trading frequency* are 0 and 0.5, indicating a significant cross-sectional variation in trading frequency. We observe that institutional investors tend to be more active traders. Mean (median) frequency of trading for institutional investors is 0.44 (0.3). In a similar vein, the average flow volatility (standard deviation of investor-level flows during the event period) for institutional versus retail clients is 3.9 and 0.78, respectively.

Even though investors in our sample do not trade too often, when they do, their trades can be large. For instance, conditional on selling, the mean and median end-investors' outflows are 50% and 38% (of their total positions), respectively. About 40% of sales are full-position sales. Among purchases, we observe that 10% of purchases are new positions. The median (mean) value of purchases that are adjustments on existing positions is 0.7% (5%). Analogous numbers for exits are 5% and 16%. Although purchases tend to be smaller than sales, they are more frequent (18% versus 82%). Overall, patterns indicate that purchases are more frequent but smaller. Sales occur less frequently but when investors sell, they sell large amounts.

Next, we assess the impact of the change in pricing structure on individual investor flows. To this end, we estimate the regression model in (3.4.4) at the investor level. Relative to equation (3.4.4), our new dependent variable is *Flow EndInv*, which is the monthly change in number of shares an investor holds in a given fund. We include investor fixed effects, which allows us to control for any permanent differences among investors and measure the differential effects due to pricing change for a given investor. We present the results in Table C.6.

As a starting point, we estimate our regression model separately for investors subjected to change (in column (1)) and those being part of the control group (in column (2)). The results indicate that investors in switching funds react less to stressed market conditions in terms of their withdrawals after the switch. On the other hand, the behavior of investors in the control group of funds that do not switch their pricing does not seem to change significantly during the same period. If anything, the effect is slightly negative, although statistically insignificant.

In column (3), we use the combined sample with the two groups of investors and estimate the relative sensitivity of the two types of investors to change using a triple-difference regression model. The results we obtain are qualitatively similar to those we obtained

from our fund-level estimation. Investors in funds with the alternative pricing withdraw relatively less of their money than do investors in traditional funds during periods of high market stress. At the same time, they put less money to alternative funds in periods of no stress.[22]

Even though the investor-level tests provide a clean identification of our economic hypothesis, our tests assume that investors differ only with respect to their time-invariant characteristics. The interpretation of the results could differ if some time-varying investor preferences (that happen to change around the switch dates) drive the differential responses of investors in pre and post periods. Given the staggered nature of switches, this is quite unlikely; moreover, the time variation in investor preferences would have to be such that investor flows are affected differently in periods of market stress versus no stress. Alternatively, investors could be affected by liquidity shocks to their entire (unobserved) wealth and rebalance their fund positions in the direction consistent with our results.

Ruling out such alternative explanations is generally challenging but we can take advantage of a unique feature in our data. Specifically, we can design tests focusing on investors who at the same time invest in funds that do and do not undergo the pricing change. Hence, any change in investor preferences is likely to be common across the two types of funds.

Since we only observe the unique investor identifiers within the same fund management company, in this test, we select the control funds from the same management company as that of treated funds. Investors in our sample do not commonly hold shares in both treated and control funds at the same time, however, we identify about 2,800 (about 10% of the sample) observations that correspond to investors with cross-fund holdings in a given month. Including investor×time fixed effects, we focus on such cases and find that the coefficient of the triple interaction term remains to be positive and statistically significant (Table C.17 of the Internet Appendix). During market stress, investors in switching funds withdraw less money than when they withdraw from traditional funds.[23]

Overall, our investor-level analysis indicates a meaningful response of the same investor within the local event window around the pricing change and provides strong evidence that alternative pricing structures affect investors' decisions and mitigate fragility in funds. The same investor is significantly less likely to redeem her shares during a stress period if the fund uses alternative pricing than if the fund uses traditional pricing.

[22]Figure C.4 shows the average differences in *Flow EndInv* between switchers (treated) and their matched funds (control) after controlling for investor fixed effects. We show differences for each event month during the $[-24, 24]$ month period. We report separate plots for periods of market stress and no stress.

[23]An additional robustness test we conduct in Table C.17 is that we rerun the tests using an alternative dependent variable, *GBP Flow EndInv*, which is the GBP monthly change in each investor's holding. Our conclusions remain similar.

3.4.4 Investment stability and alternative pricing

Our results so far suggest that open-end funds with alternative pricing structures enjoy greater flow stability, especially during market stress. In this section, we provide additional evidence to buttress this finding. First, we examine investors' flow–performance sensitivity. Second, we look at the volatility of individual investors' flows. Finally, we analyze funds' decisions to exit the market.

Flow–performance sensitivity

A well-established finding in the equity mutual fund literature is that fund flows are strongly associated with funds' performance and that the relationship between fund flows and a fund's past performance tends to be convex (e.g., Chevalier and Ellison, 1999). A recent paper by Goldstein et al. (2017) estimates flow–performance sensitivity for corporate bond funds and finds that the relationship for corporate bond funds is concave. Corporate bond funds' outflows appear to be more sensitive to bad performance than their inflows are to good performance. Goldstein et al. (2017) interpret this finding within the theoretical model of Chen et al. (2010), which predicts that the traditional pricing used by open-end funds generates strategic complementarities. If alternative pricing removes the first-mover advantage arising from the traditional pricing practice, we should expect the concavity to be lessened for swing funds. To assess this, we estimate:

$$Flow_{i,t+1} = \alpha + \beta_0\, NegAlpha \times Alternative_{i,t} + \beta_1\, NegAlpha_{i,t} + \beta_2\, Alpha \times Alternative_{i,t} + \beta_3\, Alpha_{i,t} + \beta_4\, Alternative_{i,t} + Controls_{i,t} + Time\ FE + \varepsilon_{i,d}, \quad (3.4.5)$$

where *Flow* is the flow of fund i in month $t+1$; *Alpha* is the average monthly fund alpha in the past 12 months; *NegAlpha* equals *Alpha* if alpha is below zero and it is set to zero, otherwise; Control variables are lagged *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*, all measured at month t. Following the related studies, we include year-month fixed effects to remove the time-series variation in flows and focus on exploiting the cross-fund variation in alphas. We cluster standard errors by fund and time.

Panel A of Table C.7 presents the results. In column (1), we only include *Alpha* and *Alpha*×*Alternative* to estimate differences in average flow–performance sensitivity. To evaluate any potential concavity, in column (2), we add *NegAlpha* and its interaction with *Alternative*. Consistent with Goldstein et al. (2017), we find that flows to corporate bond funds are significantly positively related to funds' past performance and this relationship is more pronounced for funds with poor performance. Most important, the results show that concavity is significantly reduced for funds with alternative pricing. In column (2),

estimated coefficients for *NegAlpha* and *NegAlpha*×*Alternative* are 5.8227 and -4.0730; both are statistically significant at the 1% level. While sensitivity to negative performance is significantly lower for funds with alternative pricing, we do not find any significant difference in sensitivity to positive performance for funds with different pricing methods. Column (3) repeats the analysis for the matched sample and confirms the robustness of these findings.

Moreover, we estimate the flow–performance sensitivity at the end-investor level using the sample of switching funds and their matching pairs. Specifically, we regress *Flow EndInv* on *NegAlpha*×*Treated*×*Post* and *Alpha*×*Treated*×*Post* while saturating the model with all other interaction terms. The analysis uses the 24-month period before and after the switch occurs. Regressions include end-investor fixed effects. Results are in Panel B of Table C.7.

Our results are reassuring and consistent with the findings obtained from the full sample. In column (1), we evaluate the overall change in the sensitivity to performance, including both positive and negative fund alphas, and we find no significant effects. In column (2), we assess the asymmetry by including interaction terms with *NegAlpha*. Similar to the full-sample results, we find significant differences in sensitivity to *NegAlpha*. Our results show that, in a switching fund, the same investor is significantly less likely to redeem her shares in the post period (*NegAlpha*×*Treated*×*Post* is -1.5247, significant at 10%). In column (3), we focus on more extreme negative performance by revising the definition of *NegAlpha* as being equal to *Alpha* when it is below the 25^{th} percentile of the sample (and zero, otherwise). Results reveal the same patterns, with amplified magnitudes—in column (3), the coefficient of *NegAlpha*×*Treated*×*Post* is -4.5641, significant at 5%.

These results provide strong evidence that alternative pricing affects only the sensitivity to poor performance. The asymmetry of the results supports the interpretation that alternative pricing mitigates the run incentives arising from traditional pricing. This is because, while there can be a *run for exit* effect on the downside, there is unlikely to be a *run to enter* effect on the upside as funds with recent good performance do not continue to perform well (e.g., Carhart, 1997; Chen et al., 2004). However, as we show in Section 3.4.6, in the absence of dilution adjustment on fund NAV, funds with poor performance experience outflows and continue performing poorly.

Volatility of end-investor flows

Another way through which fund stability may manifest is volatility of individual investors' flows. To the extent that alternative pricing reduces outflows in stress times and decreases inflows in other times, individual investors' flow volatility is expected to be lowered. To

assess this, for each investor, we calculate *Vol of Flow EndInv* (volatility of *Flow EndInv*) before and after the switch date, and estimate:

$$Vol\ of\ Flow\ EndInv_{i,t} = \alpha + \beta_0\, Treated_i \times Post_t + \beta_1\, Post_t + \beta_2\, Treated_i + \beta_3\, Controls_{i,t} + \varepsilon_{i,d}. \quad (3.4.6)$$

We present the results in Table C.18 of the Internet Appendix. The results show that, following the change in a fund's pricing, fund investors in the treatment group, in fact, have less volatile flows than investors in funds that do not undergo a change in its pricing method.

Fund exit

A direct consequence of significant fund outflows and high flow volatility is the possibility of fund liquidations. In this section, we test the role of alternative pricing methods on fund exits.

We obtain data on a fund's status from Morningstar. For each fund that exits the market, Morningstar reports the exit type: merged and liquidated. In total, we observe 40 funds (out of 299) that have exited during our sample period. Of these 40 funds, 18 are liquidated and remaining 22 are merged. On average, traditional funds are more likely to exit. About 12% (32 out of 253) of alternative funds exit, whereas this is 17% for traditional funds (8 out of 46).

We also evaluate the difference in fund exits between the two group of funds using a regression framework. The dependent variable, *Exit*, is an indicator variable that equals one if the fund exits the market during our sample period, and it equals zero if the fund is still alive.[24] *Merged* (*Liquidated*) is an indicator variable that equals one if the fund exits due to a merger (liquidation) during our sample period, and it equals zero if the fund is still alive. Main independent variable is *Alternative*. Control variables are measured in the last month before the exit occurs. Results are in Table C.8.

Columns (1)-(3) report results from the univariate regression model; in columns (4)-(6), we control for the potential impact of fund characteristics and family fixed effects; and in column (7), we report results with the matched sample. We find that, on average, alternative funds are more likely to exit but the effect is statistically insignificant. However, when we condition the sample on the type of exit, we observe a visible difference between mergers and liquidations. While we observe no consistent pattern for mergers, we find

[24] A fund exit is defined at the fund level, that is, when all of its share classes are terminated.

that alternative funds are significantly less likely to liquidate. Overall, our results indicate that alternative funds are less likely to liquidate their portfolios, arguably because they are less subject to run risks and fund flow volatility.

3.4.5 When do alternative pricing rules matter more?

Theory of runs on open-end mutual funds is linked to the presence of strategic complementarities due to first-mover advantage in the spirit of Morris and Shin (1998), Goldstein and Pauzner (2005), and Vives (2014). In this section, we exploit the variation in the strength of such complementarities across funds and investors that allows us to provide direct evidence for the mechanism described by theoretical studies.

The role of fund characteristics

Our first set of tests considers the differences between funds in terms of their fragility. We explore three hypotheses. First, the fund fragility should increase with the degree of their portfolios' illiquidity because such portfolios take longer to liquidate, and trades are more costly. We therefore expect the impact of pricing structure to matter more for funds with highly illiquid assets. Second, in the model of Chen et al. (2010), when the primary source of complementarities is the price impact of future redemptions, a large investor can internalize the negative effects of his future actions, thus weakening complementarities. Hence, fragility is likely to be higher for funds with many small investors and we expect that the pricing structure should matter more for funds with more dispersed ownership structure. In a similar vein, we expect the dispersion of ownership to be higher among funds with a large fraction of retail investors who tend to hold small shares in the fund.

To test these hypotheses, we append the specification in (3.4.2) with interaction terms, each of which capture the three dimensions of strategic complementarities. For this test, we use the full sample as the analysis requires sufficient cross-sectional variation in fund characteristics (switchers subsample is only about 10% of the full sample). We present the results in Table C.9. In column (1), we consider *Illiquidity*. In column (2), we characterize the dispersion in ownership using the Herfindahl–Hirschman Index. Specifically, *Ownership Concentration* is the Herfindahl–Hirschman Index of end-investors' ownership in a given fund. A lower value of *Ownership Concentration* indicates a more dispersed ownership. Finally, in column (3), we use $Retail = 1 - Inst$, defined as the fraction of a fund's assets held by retail investors. All specifications are based on a matched sample of funds and include a similar set of controls as before, measured as of previous month-end. Our

results support the three hypotheses we outline. The effect of alternative pricing is significantly greater for funds with more illiquid assets, funds with more dispersed ownership, and funds with more retail investors.

The role of investor characteristics

Our second set of tests exploits the variation in investor-level sensitivities. In this section, we test the two channels using unique data on individual investor types. Specifically, we link asymmetric information to investor sophistication and coordination failure to the investment horizons of investors. Institutional investors are more likely to be sophisticated (e.g., Kacperczyk and Schnabl, 2013; Schmidt et al., 2016) and therefore better understand the negative externalities stemming from NAV pricing rule. Moreover, investors with longer horizons suffer more from the dilution in fund performance due to trading costs. If alternative pricing mitigates strategic complementarities, then we would expect it to matter more for institutional investors and investors with longer horizons. To test this prediction, we estimate the regression model based on the switching experiment at the investor level. We cluster standard errors by investor and month. We present the results in Table C.10.

$$\begin{aligned} Flow\ EndInv_{i,t} = \alpha + \beta_0\, Stress_t \times Post_t \times Inv\ Type_i + \beta_1\, Stress_t \times Post_t + \\ \beta_2\, Stress_t \times Inv\ Type_i + \beta_3\, Inv\ Type_i \times Post_t + \beta_4\, Post_t + \\ \beta_5 Inv\ Type_i + \beta_6\, Stress_t + \beta_7\, Controls_{i,t} + \varepsilon_{i,d}. \end{aligned} \quad (3.4.7)$$

Inv Type is a generic variable for two measures of investor types. In columns (1) to (4), the investor type, *Inst Investor*, is an indicator variable that equals one if the end-investor is an institutional client and zero, otherwise. Column (1) shows the results for treated funds. Consistent with the strategic complementarity hypothesis, we find that institutional investors in switching funds are more likely to stay with the funds in times of stress after the fund switches to swing pricing. The behavior of institutional investors in traditional funds, in column (2), suggests no change in behavior in the post period. We perform the test of differences in the coefficients β_0 between treated and control groups and find that the corresponding p-value equals 0.07.

We extend the empirical test by asking whether a dominant type of a fund investment clientele matters for the results we document above for institutional investors. In column (3), we repeat the test considering the subsample of funds predominantly held by retail investors, and in column (4), we use funds with mostly institutional investors. The results from this analysis show that institutional investors alter their behavior more in funds with

broad retail ownership. This is in line with Chen et al. (2010) who argue that funds that are held mostly by large institutional investors can be less susceptible to strategic complementarities because large institutional investors are more likely to play a cooperative game, while coordination is harder for a more diverse retail group. Taken together with Table C.9, our analysis shows that, while the cross-fund effects are pronounced for funds with higher retail ownership, it is not driven by the retail investors that sell more in these funds, rather it is the institutional investors facing the presence of retail investors.

The investor type in columns (5) and (6) is investment horizon. Specifically, *Patient Investor* is an indicator variable that equals one if the end-investor has investing horizon above the sample median and zero, otherwise, where investing horizon is the number of months the investor holds his shares after an initial purchase.[25] In column (5), we present the results for the sample of treated funds, and in column (6) for control funds. The coefficient of the triple interaction term is positive and statistically significant for the former group, whereas it is negative, though statistically insignificant, for the latter group. The results support the hypothesis that investors with longer horizons perceive the change in pricing structure as a stabilizing force during periods of stress.

Notably, the investor type may be correlated with investment horizon. In this regard, it is useful to assess the relative contribution of the two forces to the total flow effect. In column (7), we jointly include *Stress*×*Post*×*Inst Investor* and *Stress*×*Post*×*Patient Investor*. The results indicate that both investor sophistication and investment horizon of investors in a fund are important interacting forces with the fund's pricing structure, though investor sophistication seems to be a statistically stronger factor.

3.4.6 Do alternative pricing rules affect subsequent fund performance?

A central tenet of fragility in open-end funds is that the traditional NAV-based pricing induces the dilution effect of large flows for non-transacting investors and this gives rise to first-mover advantage. A large body of empirical literature document that flow-induced trades (in particular, due to redemptions) are costly to funds and that such trades dilute fund performance (Edelen, 1999; Coval and Stafford, 2007; Alexander et al., 2007; Christoffersen et al., 2006; Goldstein et al., 2017; Feroli et al., 2014). In this section, we examine the extent to which funds with alternative pricing methods eliminate the first-mover advantage by reducing the dilution in subsequent fund performance due to investor flows.

[25] In calculating investing horizon, we use purchases before December 2014. Average horizon in traditional and alternative funds is 26 versus 30 months, respectively. This is consistent with the idea that alternative pricing rules provide protection for long-term investors.

If funds effectively use the alternative pricing rules to reduce dilution, we should expect the negative impact of investor flows on subsequent fund performance to dissipate. Moreover, the effect should be stronger for funds more illiquid portfolios, and it should be present mostly for outflows, as outflows trigger forced liquidations. In turn, inflows need not to be immediately put to force if they are to create undesired consequences. To assess this hypothesis, we estimate the following regression model:

$$AbReturn_{i,t+1} = \alpha + \beta_0\, Net\ flow_{i,t} \times Alternative_{i,t} + \beta_1\, Net\ flow_{i,t} + \beta_2\, Alternative_{i,t} + \beta_5\, Controls_{i,t} + Time\ FE + \varepsilon_{i,d}, \quad (3.4.8)$$

where *AbReturn* in month $t+1$ is calculated as the difference between a fund's return and fund's exposure to global bond market and global stock market returns. We calculate fund returns using unadjusted prices, since our focus is on the unadjusted fund performance. Fund exposures to benchmarks are calculated as $\beta 1_{t-11,t} \times Bond\ market\ return_{t+1}$ and $\beta 2_{t-11,t} \times Stock\ market\ return_{t+1}$, where $\beta 1_{t-11,t}$ and $\beta 2_{t-11,t}$ are obtained from the same 12-month rolling window regressions as alphas. Net flow includes both inflows and outflows. *Net Outflow* is the net outflow at month t, equal to *Flow* if $Flow < 0$, and to zero if $Flow \geq 0$. *Net Inflow* is the net inflow at month t, equal to *Flow* if $Flow > 0$, and to zero if $Flow \leq 0$. We cluster standard errors by fund and month.

We present the results in Table C.11. In column (1), we consider the full sample and the effect of increasing outflows on future performance. Consistent with the literature, we observe that higher outflows deteriorate subsequent fund performance for funds with traditional pricing. However, the negative impact of outflows on fund performance is almost fully eliminated for funds with alternative pricing. In column (2), we restrict our sample to funds with highly illiquid portfolios and show that the unconditional effect is amplified for such subsample. The magnitude of the effect is almost twice as large as that in the unconditional sample.

In columns (3) and (4), we present the respective results for the group of funds with inflows. The results are statistically insignificant, which corroborates our view that larger inflows may not be distortionary because fund companies have more flexibility in deploying their new capital which mitigates the associated costs. This is in line with Figure C.2 and Table C.2 that dilution adjustment factor is asymmetric in that, it is particularly high during market stress, when funds experience outflows. Overall, results in this section indicate that funds seem to be able to use the alternative pricing methods effectively enough to eliminate the dilution in fund performance arising from fund outflows.

3.4.7 The costs of alternative pricing

Our analysis shows that funds with alternative pricing mitigate redemption risk during stress periods, suggesting that these funds should be preferred over those with traditional pricing. At the same time, in the data we observe both types of funds, which suggests a potential downside to such pricing scheme. A closer inspection of our results reveals that funds with alternative pricing tend to receive less inflows outside the periods of market stress, which could rationalize the existence of the two pricing structures. In this section, we provide more detailed explanation behind the results.

We propose two plausible channels. First, the finding might reflect a possible concern among investors that fund managers' full discretion in setting adjustment factors may be detrimental to performance of their portfolios. Alternatively, the finding might be a consequence of an increase in funds' tracking errors. Funds with alternative pricing rules may arguably have higher tracking errors as these funds move their prices in response to flows which may not necessarily correspond to changes in underlying asset valuations. In our tests, we focus on the tracking error channel. Notably, while the tracking error force operates at all times, its relative detriment to investors is higher in good market conditions when the stabilizing benefits of alternative pricing are relatively smaller. Hence, the overall effect on fund flows that we observe in the data could plausibly vary over the market conditions.

We define *Tracking Error* as the R-squared obtained from the rolling 12-month factor model regressions of fund returns on global bond market and global stock market excess returns. We multiply this value by -1 so that a higher value indicates higher tracking error. Subsequently, we estimate the regression model with *Tracking Error* as the dependent variable and *Alternative* as the main independent variable. We also include the same set of controls as before. We present the results in column (1) of Table C.12.

Consistent with our hypothesis, we find a positive and statistically significant coefficient of *Alternative*, which means that funds with alternative pricing generate higher tracking errors. The relevant question is whether the higher error inhibits the level of flows these funds receive. In column (2), we provide a benchmark specification of fund flows conditional on the pricing structure. The coefficient of *Alternative* is negative but borderline insignificant. In column (3), we additionally include *Tracking Error* as an explanatory variable. The results indicate that tracking error is in fact an important determinant of fund flows; the coefficient of *Tracking Error* is negative and statistically highly significant. At the same time, once we include *Tracking Error* in the regression, the coefficient of *Alternative* becomes nearly zero, suggesting that an important part of its negative effect on flows is captured by differences in tracking error.

As a final test, we evaluate whether the negative effect of tracking error diminishes the growth of investor base in funds with alternative pricing in good market conditions, which is when the effect of tracking error is particularly costly. To this end, we define *New Investor*, which is the number of new investors entering a fund in a given month divided by the funds total number of investors as of previous month-end. We present the results from estimating the regression of *New Investor* on *Alternative*. Column (4) shows that funds with alternative pricing are in fact able to attract significantly fewer new investors outside periods of high market stress.

3.4.8 Do fund companies internalize their investors' decisions?

Given that a fund's pricing structure is a way to address a problem of possible fund fragility, in this section, we examine whether funds treat their pricing scheme as a substitute for other hedging instruments. Based on the literature, we examine three options: increased cash holdings, reduced asset concentration, and fund load fees.

We define cash (*Cash*) as a fund's total cash holdings (including cash equivalents) divided by the fund's total assets. Asset concentration (*Asset Conc*) is Herfindahl–Hirschman Index of a fund's asset holdings in each month. *Front Load* is the value-weighted average of (minimum) front load charges across share classes of a given fund; *Back Load* is the value-weighted average of (minimum) back load charges across share classes of a given fund. We assess the relationship between pricing structure and the alternative hedging instruments by estimating the following regression model:

$$Hedging\ Instrument_{i,t+1} = \alpha + \beta_0\, Alternative_{i,t} + \beta_1\, Controls_{i,t} + Time\ FE + \varepsilon_{i,d}, \tag{3.4.9}$$

where *Hedging Instrument* is a generic name for different instruments. We present the results in Table C.13. We find that funds with alternative pricing rules hold less cash, on average, consistent with the hypothesis that cash and alternative pricing rule are substitutes for each other. On the other hand, the coefficients for asset concentration, front load, and back load, though negative, are all statistically insignificant. The results are consistent with those of Chen et al. (2010) who argue that loads cannot eliminate dilution in fund value due to flows, therefore they cannot address strategic complementarities. This is because proceeds from loads are not retained in the fund; rather loads are usually paid out to the distribution channel.

3.5 Conclusion

Open-end mutual funds globally manage tens of trillion of dollars in assets. Quite often, these assets are illiquid making the conversion to liquid assets difficult, especially at times of significant market stress. Such liquidity mismatch in combination with strategic complementarities arising from NAV-based pricing rule pose a significant threat to these companies and the broad market. Mitigating the fragility risk is of first-order importance to financial institutions managing these companies, their investors, and policy makers concerned with financial stability and social welfare. In this paper, we empirically evaluate the effectiveness of one tool, swing pricing, that allows for dilution adjustment of funds' net asset values.

Through the FCA, we obtain detailed data on UK-based corporate bond mutual funds with different pricing practices and investors' base. Using a combination of micro-level identification strategies that address endogeneity concerns, we show that alternative pricing rules change open-end funds' operations in a way that enables funds to more effectively manage their liquidity risk. They reduce the degree of redemptions during periods of high market stress. The stabilizing effect is particularly visible for institutional investors and investors with long investment horizons, which supports the presence of strategic complementarities in the data. Moreover, these funds eliminate the first mover advantage by removing the dilution in fund performance arising from fund outflows.

Although the results indicate that swing pricing may be a useful financial stability tool, our analysis also documents an important cost associated with such rules: funds with alternative pricing rules have difficulty attracting new investor capital outside the crisis periods, largely because their portfolios exhibit greater tracking errors. The clear dominance of alternative pricing structure is therefore difficult to establish. Future research can aim to evaluate swing pricing relative to its potential alternatives.

Our results offer important policy implications. Recently, policy makers have expressed concerns with the growing illiquidity mismatch in various parts of the asset management industry. For example, in his policy speech on June 26, 2019, the Governor of the Bank of England, Mark Carney, called for actions preventing possible systemic runs on the industry arising from significant illiquidity mismatch. However, such concerns may be muted if managers efficiently use available pricing rules. Simultaneously, regulation permitting alternative pricing rules has become effective in the U.S. only recently, in November 2018. Given similarity in investor types and general development of European and U.S. markets, our results may help to understand the expected effects of the new regulation for the U.S. market.

Chapter 4

The Drivers and Inhibitors of Factor Investing

Abstract: I model the equilibrium asset allocations when households can invest directly, search for factor (smart-beta and ETFs) investments or fundamental (stock-picking) investments. Managers endogenously choose to specialize in factor or fundamental information given the equilibrium fee structure. Fundamental managers can opt to be opportunistic "closet indexers." I show that wealth *inequality* increases demand for factor investing: fundamental investing attracts the wealthiest households, who are more willing to detect closet indexing. Fundamental managers have to compete more aggressively through information acquisition, which lowers their excess returns and thus delegation fees. The reduced fundamental fees force factor managers to lower their fees, making factor investing even more attractive. However, the equilibrium fraction of capital allocated to factor investing can never reach 100 percent: the ceiling is determined by the endogenous level of opportunism in the fundamental investment industry.

4.1 Introduction

In August 2019, the net asset value of U.S. indexed funds reached $4.27 trillion, surpassing for the first time ever the money managed by their stock-picking rivals.[1] A large proportion of this increase in passive allocation is explained by factor investing: over $1.9 trillion is currently invested in factor funds, including index-based vehicles, rules-based ETFs, and smart-beta products; and this number is expected to swell to $3.4 trillion in 2022.[2]

[1] 'Index Funds Are the New Kings of Wall Street,' the Wall Street Journal, September 2019. Figure D.1 in the appendix summarizes the facts.

[2] 'Can Factor Investing Kill Off the Hedge Fund?' Financial Times, July 2018; 'Hybrid Funds Smooth Path Between Active and Passive Strategies,' Financial Times, November 2019.

Factor investing allows households to buy market beta and hold broadly diversified portfolios with minimal delegation fees. In contrast, traditional active funds exploit differences between the market price and the fundamental value of individual stocks through rigorous security selection. As factor investors partly free-ride on the price discovery efforts of fundamental investors, the rise of factor investing might distort asset prices and affect price efficiency.[3] This raises a number of questions: What causes the increase in factor investing? If the rise is problematic, is there an upper bound of the fraction of capital that will ultimately be allocated to factor investing?

In this paper, I address the questions through a rational expectations equilibrium (REE) model, which jointly determines the (i) equilibrium asset prices, (ii) households' allocation between factor, fundamental, and direct investing, (iii) managers' selection into different roles, and (iv) delegation fees. Investment in acquiring fundamental information is costly and unobservable. Thus active managers can opt to become "closet" indexers, who opportunistically market themselves as fundamentally informed, despite holding index-like portfolios.[4] Analysis of the equilibrium reveals that there exists an implicit *ceiling* for factor investing: as factor investing increases, more and more of the market-level information is impounded into asset prices. This effect makes the return differential between fundamental investing and factor investing larger, and thus ensures that factor investing will never completely dominate fundamental investing. The upper bound of factor investing is lower when it is harder for managers to obtain fundamental information, or when it is easier for households to detect closet indexing.

A novel feature of this work is to incorporate elements of household finance into the standard REE model. I introduce heterogeneity among households in the form of wealth dispersion and examine how this affects capital allocation in the fund industry.[5] I find that *wealth inequality* fuels demand for factor investing and leads to compression of delegation fees. Fundamental investing attracts the wealthiest households, who are most willing to detect closet indexing. Thus active managers have to increase information acquisition, curtailing the marginal gain from information production. This effect brings market prices closer to their fundamentals, endogenously generating decreasing returns to scale of the fundamental investment industry and forcing active funds to lower fees to retain capital. Factor funds, accordingly, charge less and gain popularity.

[3]Empirical evidence can be found in Coles et al. (2018a) and Israeli et al. (2017).

[4]The terminology "closet" fund first appears in Oppel (1999), and is widely used in the later empirical studies, e.g., Petajisto (2013) and Cremers et al. (2016). Cremers and Petajisto (2009) find that internationally, about 20% of the mutual fund assets are managed by closet funds.

[5]None of the existing literature has taken into account how household inequality is linked to capital allocation in the fund industry, though there exist some studies examining the effect of household heterogeneity on asset prices (cf. Peress, 2003; Gomes and Michaelides, 2007).

To derive these ideas, I extend Grossman and Stiglitz (1980) model to a multi-asset general equilibrium. The market is not completely efficient and is subject to exogenous and noisy supply of assets. Managers are incentivized and endogenously choose to specialize in factor or fundamental information. Households can detect the true skill of active managers through a costly search, or invest based on their prior information. Compared to factor investing, fundamental investing is associated with higher active fees, and an implicit cost arising from opportunism. The delegation fees are endogenously set. The model shows that factor investing increases competitive pressure, forcing active funds to lower fees, thereby providing theoretical justification for the finding of fee compression in Cremers et al. (2016).

The equilibrium capital allocation of households depends on security market characteristics, risk preferences, information production costs, and overall market efficiency. Generally, factor investing is more popular than fundamental investing due to its efficient cost and the endogenous opportunism in the active industry. However, when the costs of obtaining fundamental information are sufficiently low, or when the gains from trading on idiosyncratic risks are sufficiently large, factor investing will be driven out of the market. Above a certain threshold, factor strategies suffer a severe reduction in the diversification benefit, as the constituents in the index comove a lot and index price becomes highly informative, driving households to fundamental investing.

To see if capital allocation and other equilibrium characteristics persist in a more realistic context, I introduce household heterogeneity to the baseline model in two dimensions: wealth inequality and heterogeneous cognitive ability of finding and vetting a manager. Household heterogeneity has the following two implications: (i) while all three investment vehicles coexist, at the individual level households have different preferences with regard to their capital allocation; and (ii) the allocation strategies from individuals with different wealth positions, thus, contribute differently to funds' inflows. First, heterogeneity yields insights into the likely composition of fundamental investing households: those with higher wealth endowments or more insight into the quality of funds, favor fundamental investing. In the case of cognitive ability, more able households can more easily discern fund manager opportunism; in the case of wealth, wealthier households exhibit less absolute risk aversion at the margin.

Second, the model predicts that higher inequality fuels demand for factor investing, mainly at the expense of direct investing. The wealth effect is more prominent than the cognitive ability effect. Hence, household inequality serves as one possible explanation for rising indexation, as both time trends can be observed simultaneously in countries such as the U.S.[6] Increased inequality concentrates wealth in the hands of less risk-averse wealthy

[6]See a time-series summary of wealth inequality in Figure D.2, which lines up with the time trend in Figure D.1.

households, who favor fundamental investing and have relatively low searching cost to detect closet indexing. This increases the attractiveness of managers becoming skilled, leading to higher inside competition. Thus the outperformance of fundamental strategy shrinks, making households less interested in fundamental investing. Moreover, the reduced fees earned by active funds force factor funds to lower their fees, making factor investing even more attractive compared to direct investment.

Third, in a heterogeneous equilibrium, market competition can never completely eliminate closet indexers. For instance, when fundamental information production is rather costly, active managers charge a high premium and are required to generate large excess return to justify this charge. However, the less wealthy or less able households can never be incentivized to fully invest with skilled managers, limiting the gains from fund inflows through skill acquisition. In contrast, in the homogeneous world, households are *ex-ante* identical and can exert effort to learn about managerial abilities. Facing the high cost of active delegation, *ceteris paribus*, households are more likely to become informed to avoid selection error. As a consequence, active funds have to be truly skilled to secure inflows, and closet indexers are wiped out.

This paper provides a new lens for understanding the drivers of the rise in factor investing. I find that increased factor investing can be attributed to the following factors: (i) A falling but not insignificant searching cost for households: widespread access to information online makes it easier for households to learn about managerial abilities through assessing managers' investment principles, trading infrastructure, etc. A decrease in searching cost lowers the excess returns generated by fundamental investors, and thus spurs more factor strategies.[7] (ii) A falling but not degenerate fundamental information cost for managers: this is highly possible due to more transparency in the corporate information environment and more developed techniques like data mining. Lower information production cost forces both funds to charge less, making professional delegation more attractive. (iii) An intermediate level of risk aversion: households are neither so afraid of risk that they prefer riskless assets, nor so risk-inclined that they ignore the volatility of undiversified portfolios.[8] (iv) Higher idiosyncratic risk: if the idiosyncratic factors are volatile or their supply is highly uncertain, risk-averse households seek factor investing as a substitute. (v) Increased inequality: as explained above, more efficient individual prices and fiercer competition within the active industry encourage factor investing. These findings help

[7]In the context of my model, I find that a sufficiently small searching cost engenders a significantly low level of opportunism, and thus leads households to shift towards fundamental investing and away from factor investing. However, it is highly unlikely that the searching costs are negligible in a real-world context.

[8]This non-monotonic effect of risk appetite contradicts the argument in Glasserman and Mamaysky (2018).

understand a number of empirical observations. The comparative statics have further implications on return volatility, co-movement, and price efficiency.

Related literature

The paper contributes to an expanding and diverse body of literature. There exists massive literature modeling delegated portfolio management. Some introduce household behavioral biases, such as what is termed "trust" in Gennaioli et al. (2015). Others take rational expectations assumption built on the Grossman and Stiglitz (1980) framework. Among these REE studies, Gârleanu and Pedersen (2018) model the investors searching for active managers. However, their model is restricted to single-asset setup. Kacperczyk et al. (2016) take the attention allocation exogenous and in their framework all informed agents ultimately form identical strategy. In closely-related contemporaneous work, Glasserman and Mamaysky (2018) and Gârleanu and Pedersen (2019) endogenize the information choice of agents and prove Samuelson's dictum. This work is in contrast with theirs, in that it delivers distinct prediction on the drivers of factor investing, and explicitly models the opportunistic closet funds. None of these studies address the link between household heterogeneity and their delegation choices.

The other way of deriving the equilibrium is through the Kyle (1985) framework. For instance, Cong and Xu (2016) examine the rising factor investing from the security design perspective, on how it affects trading behaviors and asset prices. Bhattacharya and O'Hara (2018) borrow a similar setting and document that ETF activities introduce fragility and alter the information efficiency of underlying markets via herding. But they both introduce indexing activity as exogenous, which, in my model, instead, is one of the pertinent features of asset market equilibrium.

Empirically, Sorensen et al. (1998) discuss the allocation trade-off between active and index funds, with the optimal strategy of an asset owner largely depending on the stock-picking skills of active funds. In fact, even though on average, active industry does not outperform indexation, a bunch of literature documents that a modest amount active managers possess "skills" to generate excess return, including both stock-picking and market-timing.[9] The allocation between different investment vehicles may have further implications on fund structure and fees, as what is emphasized in this paper. Vayanos and Woolley (2016) raise the concern that adopting index-based strategy can circumvent the agency problem. International evidence has been found by Cremers et al. (2016), that increased indexing forces the active rivals to lower fees.

[9]For reference see Kosowski et al. (2006); Kacperczyk and Seru (2007); Mamaysky et al. (2008); Cuthbertson et al. (2008); Cremers and Petajisto (2009); Kacperczyk et al. (2014); Berk and Van Binsbergen (2015); Choi et al. (2016); Pástor et al. (2017).

To the best of my knowledge, this paper is the first to combine the standard REE model with household finance, and yield clear prediction of household heterogeneity affecting delegation preference and asset market. Wealth inequality and asymmetric cognitive ability among households, nevertheless, have long been crucial social-economic topics. High wealth inequality implies low future stock returns, as found by Gomez et al. (2019). Concerning cognitive ability, Grinblatt et al. (2011) examine its relation with stock market participation. Benjamin et al. (2013) conduct a laboratory study which links cognitive capability to risk preference. Some economic literature examines similar questions in a different background, including Proto et al. (2019) that design an experiment on how heterogeneous intelligence level affects outcomes of groups.

The empirical literature towards the effect of indexing on asset price efficiency is growing, though a bit mixed. Israeli et al. (2017) find that over the long run ETF activities deteriorate price efficiency. However, Glosten et al. (2016) argue that ETFs improve especially price efficiency for stocks with weak information environments or low liquidity. The amplification of factor investing also affects return volatility and co-movement of assets.[10]

The remainder of the paper is structured as follows. Section 4.2 describes the economic setup. I derive the asset market equilibria in Section 4.3. Section 4.4 introduces heterogeneity of households to the standard REE model. Section 4.5 summarizes the possible drivers of rising factor investing, and uses them to interpret other pertinent features of asset market equilibria. Section 4.6 concludes. All proofs are in the appendix.

4.2 Economic Setup

4.2.1 Assets

Assume there are $l+1$ underlying assets in the market. One is the riskless asset with gross return normalized to be 1. The other l risky assets deliver the payoffs $\boldsymbol{v} = (v_1, v_2, ..., v_l)^\intercal$, where

$$v_i = \beta_i S + \mathcal{E}_i, \qquad i = 1, 2, ..., l. \tag{4.2.1}$$

Intuitively, each asset can be decomposed to include a *factor loading*, S (measuring systematic risk shared by all publicly traded assets), and an *idiosyncratic loading*, $\mathcal{E}_i$ (measuring, for instance, how a firm-specific shock affects stock value). S and $\mathcal{E}_i$ are independent and normally distributed, with $S \overset{d}{\sim} \mathcal{N}(\overline{s}, \sigma_s^2)$ and $\mathcal{E}_i \overset{d}{\sim} \mathcal{N}(0, \sigma_e^2)$, $i \in \{1, 2, ..., l\}$. β_i captures

[10] There exist empirical discussions from academia (e.g., Glosten et al., 2016; Israeli et al., 2017; Da and Shive, 2018; Ben-David et al., 2018; Coles et al., 2018a) as well as from regulation and law authorities (e.g., Sushko and Turner, 2018; Anadu et al., 2018; Rock and Kahan, 2018; James et al., 2019).

asset i's exposure to the aggregate market. Any portfolio held by an investor can be decomposed to an l-dimensional vector, representing the holding of the risky assets, denoted by $\boldsymbol{x} \in \mathbb{R}^l$. The holding of riskless asset is thus automatically determined.

The index portfolio is a bundle of underlying assets, with a composition of $\boldsymbol{x}_I = (b_1, b_2, ..., b_l)^\intercal$, where $\sum_i^l b_i = 1$. Its payoff is $\sum_i^l b_i v_i = S$, i.e., corresponding to the market portfolio. This implies that (i) $\sum_i^l b_i \beta_i = 1$ and (ii) $\sum_i^l b_i \mathcal{E}_i = 0$. Condition (ii) can be guaranteed by the assumption that, within the index portfolio, idiosyncratic risks are averaged out, i.e.,

$$Cov(\mathcal{E}_i, S) = Cov(\mathcal{E}_i, \sum_{k=1}^{l} b_k \mathcal{E}_k) = 0, \quad \forall i \in \{1, 2, ..., l\}.$$

These l equations imply a feasible $l \times l$ correlation matrix of the asset-idiosyncratic loadings that sustains condition (ii).

Assume an aggregate market supply, $\boldsymbol{q} \in \mathbb{R}^l$, which is random and satisfies a jointly normal distribution, $\mathcal{N}(\overline{\boldsymbol{q}}, \Sigma_q)$. Random variation captures the existence of a group of "noise traders," or, as in the literature, termed "liquidity traders." Noisy trading is partly attributed to the market activities including initial public offering, stock repurchase, float adjustment by share control groups, etc. Noise traders buy an exogenous amount of shares of the assets, $\boldsymbol{q} - \overline{\boldsymbol{q}}$.

4.2.2 Households, managers, and information

Consider an economy where households can invest either directly or through professional delegation, typically, through funds. Thus, the economy is populated by three types of market participants: direct investors, factor investors, and fundamental investors. Factor strategies, deployed by factor funds, place bets on market factor (driven by, for example, macroeconomic announcements and overall market turbulence), and adjust exposure based on predetermined financial metrics. Fundamental investing is implemented by active equity managers, who generate excess returns from anomalies in individual security prices through informed analysis and active stock selection. Compared to fundamental investment, factor investment amounts to passive allocation, as factor funds cannot tilt their portfolios towards one particular asset.[11]

The incentives of informed trading underpin the rationale that the market is mostly, but not completely, efficient. All investors are risk averse and price takers, and, if endowed with the same information package, can be interpreted as a continuum of identical investors. I will extend the model to incorporate heterogeneous investors in Section 4.4.

[11] Existing literature that motivates indexing in this fashion includes Cong and Xu (2016), Bond and Garcıa (2017), and Glasserman and Mamaysky (2018).

The certainty-equivalent wealth of households investing directly, or through factor or active funds, denoted by $u_j(\cdot)$ for $j \in \{h, b, a\}$ respectively, can be separated from their initial wealth W:

$$u_j(W) = -\frac{1}{\gamma} \log\left(\mathbb{E}\left[\max_{\mathbf{x}_j} \mathbb{E}\Big(e^{-\gamma \left(W + \mathbf{x}_j^\intercal (\boldsymbol{v} - \boldsymbol{p})\right)} \Big) \Big| \mathcal{I}_j \right] \right) = W + u_j(0) \equiv W + u_j, \quad (4.2.2)$$

where $\boldsymbol{p} = (p_1, p_2, ..., p_l)^\intercal$ is the price vector, which will be internalized in equilibrium. I assume an identical risk-averse coefficient across all investors, $\gamma > 0$. For investor type j, $\mathcal{I}_j$ is his information set and $\boldsymbol{x}_j$ is his portfolio holding strategy, specified later in this section.

Managers

There exist imperfect signals in the market. Factor investors observe a signal of systematic risk S:

$$s = S + \varepsilon_0,$$

where S is independent of ε_0, $\varepsilon_0 \overset{d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. Note that $\sigma_\varepsilon^2 = 0$ corresponds to perfect information and $\sigma_\varepsilon^2 = +\infty$ corresponds to no information. The price of the index portfolio is denoted by p_I, which, by no arbitrage condition, satisfies $p_I = \sum_{i=1}^{l} b_i \, p_i$, i.e., equal to the weighted sum of the prices of underlying stocks. Assume the total mass of factor funds is one.

Fundamental investors acquire asset-specific information and make investment decision at the individual stock level. Thus, besides factor loading, they also have imperfect inside information concerning idiosyncratic risks, $\boldsymbol{e} = (e_1, e_2, ..., e_l)^\intercal$, with

$$e_i = \mathcal{E}_i + \varepsilon_i, \qquad i = 1, 2, ..., l,$$

where $\mathcal{E}_i$ is independent of ε_i, all ε_i are mutually independent and satisfy $\varepsilon_i \overset{d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. However, as in Gârleanu and Pedersen (2019), skill acquisition is unobservable. Thus there exist "closet" funds, who opportunistically market themselves and charge fees as fundamental investors, but simply trade on overall market signals like factor investors. Assume active funds are also a continuum with mass one, among which a fraction $\mu \in [0, 1]$ is skilled, and the rest are opportunistic indexers. In an equilibrium, an endogenous amount of active managers choose to acquire skills at a cost of investigation $k > 0$.

Households

A mass λ_b of households choose factor investing, and a mass λ_a go into active funds. The rest of households, with mass λ_h, self-invest in the riskless bond and risky stocks based on public signals.

Households who favor fundamental investing are *ex-ante* unaware of which active fund manager is skilled. Thus, according to the setup, they have a probability μ of correctly picking the skilled ones, and a probability $1-\mu$ of misallocation. But they can incur some searching cost (e.g. in information acquisition and due diligence), c, in ascertaining the manager quality. Assume there is $\xi \in [0,1]$ fraction of households who become costly informed so that they can identify the skilled managers. The rest $1-\xi$ households invest based on prior information. In the equilibrium, household investment in finding out manager quality is endogenous.

To sum it up, the information set of direct investors, factor investors (including closet indexers), and fundamental investors is, respectively,

$$\mathcal{I}_h = \{\boldsymbol{p}\}, \quad \mathcal{I}_b = \{\boldsymbol{p}, p_I, s\}, \quad \mathcal{I}_a = \{\boldsymbol{p}, s, \boldsymbol{e}\}. \tag{4.2.3}$$

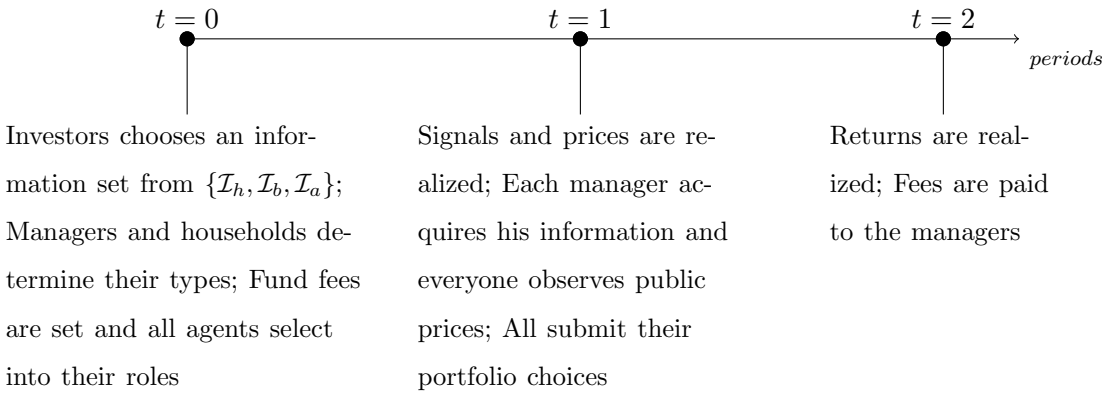
The amount of households allocating money to these three investment vehicles are, correspondingly,

$$\lambda_h, \quad \lambda_b + (1-\mu)(1-\xi)\lambda_a, \quad \big(\xi + \mu(1-\xi)\big)\lambda_a.$$

All investors fall into the three categories. Mathematically I assume

$$\lambda_h + \lambda_b + \lambda_a = 1.$$

In an equilibrium: (i) households endogenously allocate their capital between the three investment vehicles; (ii) fund managers endogenously choose to specialize in factor or fundamental information; (iii) households, if accessing the market through active funds, endogenously invest in finding out manager quality; and (iv) active managers self-select between being skilled and opportunistic. The delegation fees are endogenous and are set through Bertrand competition. The sequence of events is summarized below.



4.2.3 Portfolio holding strategies

Each group of investors picks his portfolio holding $\boldsymbol{x}_j$ ($j \in \{h, b, a\}$) to maximize his utility (4.2.2). According to the standard portfolio theory, investors pick the mean-variance efficient portfolio based on their choice of products.

Direct investors only observe the public prices and choose their portfolio correspondingly. For tractability, I assume that direct investors, in aggregation, hold the portfolio equivalent to the market portfolio. It can either because (i) they lack the expertise and skills thus, all replicate the market portfolio, or (ii) they all invest in part of the assets belonging to the index due to accessibility and aggregate to the market portfolio. Thus direct investors deploy a strategy equivalent to $\boldsymbol{x}_h = \kappa_h \boldsymbol{x}_I$, where $\kappa_h \coloneqq \frac{1}{\gamma} \frac{\mathbb{E}[S|p_I] - p_I}{Var(S|p_I)}$. Factor investors profit from timing the market factor. Thus their strategy is defined by $\boldsymbol{x}_b = \kappa_b \boldsymbol{x}_I$, with the proportion $\kappa_b \coloneqq \frac{1}{\gamma} \frac{\mathbb{E}[S|s,p_I] - p_I}{Var(S|s,p_I)}$. Fundamental investors actively pick their portfolio components and thus hold the following,

$$\boldsymbol{x}_a = \frac{1}{\gamma} \boldsymbol{\Sigma}_a^{-1} \big(\mathbb{E}[\boldsymbol{v}|\mathcal{I}_a] - \boldsymbol{p} \big), \tag{4.2.4}$$

where $\boldsymbol{\Sigma}_a$ is the variance-covariance matrix conditional on their information set, and $\mathbb{E}[\boldsymbol{v}|\mathcal{I}_a]$ is the conditional mean payoff of stocks.

Because the risky stocks are correlated both in terms of systematic loadings and idiosyncratic loadings, it becomes tricky to calculate the conditional variance-covariance matrix. Therefore I apply the Gram-Schmidt *orthogonalization* method, with details in Appendix Section D.2.

4.2.4 Equilibrium price and price efficiency

The market clearing condition is

$$\lambda_h \boldsymbol{x}_h + \big(\lambda_b + (1-\mu)(1-\xi)\lambda_a\big)\boldsymbol{x}_b + \big(\xi + \mu(1-\xi)\big)\lambda_a \boldsymbol{x}_a = \boldsymbol{q}. \tag{4.2.5}$$

Because investors' portfolio holding strategies are functions of prices, from the demand-supply relation (4.2.5) one can figure out the equilibrium asset prices, as given by Proposition 4.2.1.

Proposition 4.2.1 (**Equilibrium prices**) *The equilibrium prices for factor loading (p_I) and idiosyncratic loadings ($\tilde{p}_i,\ i = 1, 2, ..., l$), are linear in supplies and signals:*

$$\begin{aligned}
&p_I = \theta_0 + \theta_b\big[(s-\overline{s}) - \theta_q(q_I - \overline{q}_I)\big],\\
&\tilde{p}_2 = p_2 - \beta_2 p_I = \phi_a\big[(e_2 - 0) - \phi_q(q_2 - \overline{q}_2)\big],\\
&\tilde{p}_3 = p_3 - \beta_3 p_I = \psi_a\big[(e_3 - 0) - \psi_q\big((q_3 - \overline{q}_3) + \psi_\rho(q_2 - \overline{q}_2)\big)\big],\\
&\dots
\end{aligned}$$

where $q_I \overset{d}{\sim} \mathcal{N}(\overline{q}_I, \sigma^2_{q_I})$ is the supply of S, $q_2 \overset{d}{\sim} \mathcal{N}(\overline{q}_2, \sigma^2_{q_2})$ is the supply of $\mathcal{E}_2$, and $q_i \overset{d}{\sim} \mathcal{N}(\overline{q}_i, \sigma^2_{q_i})$ for $i \geq 3$ are the rest orthogonal supplies specified in Appendix Section D.2. $(\theta_0, \theta_b, \theta_q)$, (ϕ_a, ϕ_q), and $(\psi_a, \psi_q, \psi_\rho)$ are parameters determined in the proof.

With *no arbitrage* property one can easily recover the price vector $\boldsymbol{p}$. In the proof, I first assume a linear form, and determine the unique value of the coefficients. Then by construction the linear price vector is the unique equilibrium price. The proof shows that, in the equilibrium, any asset component has a weight in the index proportional to its expected market capitalization. When there is a positive supply shock, the prices of both the index and individual assets fall. Thus, an additional supply will make the assets more attractive to both uninformed and informed investors.

I inherit the canonical measure of price *efficiency* from Grossman and Stiglitz (1980). Hence, the informativeness of index price is $\frac{Var(S|s)}{Var(S|p_I)}$, corresponding to the precision of information signal about the innovation to the fundamental contained in asset price. The number falls in $(0, 1)$ as signal contains more information compared to price thus reduces variation. For computational convenience, I concentrate on the quantity

$$\eta_I \coloneqq -\frac{1}{2}\log\Big(\frac{Var(S|s)}{Var(S|p_I)}\Big) \in (0, \infty), \tag{4.2.6}$$

termed as price *inefficiency*. Index price becomes more efficient when there are more

informed trades of market factor, represented by a smaller η_I. Note that idiosyncratic information, in aggregation, does not affect the informativeness of the price of market portfolio.

Similar to index price inefficiency (4.2.6), I define price inefficiency for individual asset i,

$$\eta_i \coloneqq -\frac{1}{2}\log\left(\frac{Var(v_i|\mathcal{I}_a)}{Var(v_i|p_i)}\right) \in (0,\infty), \quad \forall i \in \{1,2,...,l\}. \tag{4.2.7}$$

The informativeness of price can be directly linked to the relative performance of funds (see Appendix Section D.3.2). How do exogenous changes affect price efficiency will be answered after characterizing the equilibrium.

4.3 Asset Market Equilibrium

I have characterized the equilibrium asset price and its informativeness in the previous section. In this section, I develop the full asset market equilibrium, with the coexistence of (i) direct investors and fund users, (ii) both factor, opportunistic and skilled active managers, and (iii) households that ascertain the manager type through searching and that invest based on prior information. Without loss of generality, I focus on a model with two risky assets ($l = 2$) for tractability. Nevertheless, the rationale of the analyses prevails regardless of the number of assets. Especially, in Appendix Section D.5, I show that the results are applicable to a market with a large amount of assets.

The asset market equilibria will be derived following several steps. First, in Section 4.3.1, I derive the equilibrium fees and net-of-fee performance of funds, based on the competitive market assumption.[12] Second, I discuss the entry of skilled active managers and informed households in Section 4.3.2. Third, by internalizing the amount of factor investing and fundamental investing, I obtain the main results of the interior equilibrium as well as of several boundary equilibria, as presented in Section 4.3.3. Fourth, the estimated equilibrium conditions imply an upper bound of the fraction of capital in factor investing. The intuition is provided in Section 4.3.4.

4.3.1 Fees and performance

Funds charge fees for delegated portfolio management. These fees, in the real market, are usually a fixed fraction of the fund's assets under management, conventionally termed

[12]This is in contrast to Gârleanu and Pedersen (2018), where the authors rely on Nash bargaining to determine fees.

"management fees." Fees, in my model, will be set through Bertrand competition.

A household who delegates her wealth to factor funds, with initial wealth W, has to pay a fee α_b. Her net-of-fee utility is $W + u_b - \alpha_b$.[13] Whereas if no manager is employed, she has an outside option yielding a utility of $W + u_h$. Hence a household's net gain from factor investing is independent of her initial wealth level W. So do households choosing fundamental investing.

Similarly I can determine the fees of active managers. All active funds, skilled or opportunistic, charge an identical fee α_a, because their types are *ex-ante* unrevealed to the public. This implies that part of the managers are unfairly earning a compensation package from hidden information. Households are thus incentivized to find out about managerial capabilities.

An informed household weakly prefers fundamental investing to self investing if and only if

$$u_a - c - \alpha_a \geq u_h.$$

In an equilibrium, the informed households, after exerting cost c, are net indifferent between investing directly and with active funds, and are able to avoid closet indexers. Similar coexistence implies that households are (i) indifferent between fundamental and direct investing, (ii) indifferent between factor and direct investing, and (iii) indifferent between incurring searching cost to become informed or staying uninformed. The existence of four investment vehicles (direct, factor, closet, and skilled active investment) guarantees the following three equalities of utilities defined in (4.2.2),

$$u_a - c - \alpha_a = u_h, \tag{4.3.8}$$

$$u_b - \alpha_b = u_h, \tag{4.3.9}$$

$$\mu\, e^{-\gamma\,(u_a - \alpha_a)} + (1-\mu)\, e^{-\gamma\,(u_b - \alpha_a)} = e^{-\gamma\,(u_a - \alpha_a - c)}. \tag{4.3.10}$$

One can see from the above expressions that a higher gross return of fund, u_j, guarantees more monetary fee payments to the j type of managers ($j = a, b$), which captures in a stylized fashion the outperforming incentive arising from the fractional fees.

To an uninformed yet risk-averse household, investing in active funds gives rise to an additional fund-selection risk. This risk is especially high when only a small fraction of managers are skilled. Thus, the expected utility from uninformed investing is not just the

[13] If there is any concern regarding why not setting fees proportional to the realized payoff, I show here that the monetary and fractional setups are identical, yet the former greatly simplifies calculation. Suppose funds charge fractional fees $\hat{\alpha}_j$ ($j = b, a$), then the excess output to the households after fees are $(W + u_j)(1 - \hat{\alpha}_j) - (W + u_h) \approx (1 - \hat{\alpha}_j)\, u_j - u_h$. It holds approximately because the delegation fees of mutual funds are usually small. Then simply replacing $\alpha_j = \hat{\alpha}_j\, u_j$ in a framework of fractional fees, all the conclusions in my model still hold.

weighted average payoffs of picking the skilled and opportunistic managers, but with an extra premium compensating the risk of random selection. The concave combination in condition (4.3.10) reveals this risk premium implicitly. This is in essence identical to Berk and Green (2004), that managers on average deliver zero outperformance after fees. In this paper, however, I allow outperformance or underperformance at the individual manager level. The fact that active funds outperform the index benchmark before fees, yet not after fees, is well-documented in the empirical literature.[14]

Skilled active managers spend cost k to acquire "skill." They subsequently benefit from attracting more capital, as they will gain money from informed individuals with certainty. Closet and factor fund managers' costs of acquiring information and formulating portfolios are normalized to zero. The fixed information costs for households, c, and for managers, k, are both exogenous. Similar to Brown and Davies (2017), I assume the fees collected by active managers are decomposed into two components: (i) a fee for the asset management service (coin the fee for factor funds), and (ii) a fee for active stock selection. Mathematically,

$$\alpha_a = \alpha_b + k. \tag{4.3.11}$$

Hence from the above analysis, I present the following results regarding the fee structure and performance of different types of funds, given price informativeness.

Lemma 4.3.1 (**Fees and performance**) *The delegation fees of factor investing and fundamental investing are, respectively,*

$$\alpha_b = \frac{\eta_I}{\gamma}, \qquad \alpha_a = \frac{\eta_I}{\gamma} + k. \tag{4.3.12}$$

In equilibria:

(i) Before fees, on average, active funds outperform factor investing, i.e., $\mu\, u_a + (1 - \mu)\, u_b > u_b$*;*

(ii) After fees, the outperformance of active asset management disappears, typically for risk-averse households.

Moreover, the equilibrium fees, combined with condition (4.3.8), imply the following relationship,

$$\alpha_a = \alpha_b - c + (u_a - u_b).$$

This provides some explanation for the fact that, some active funds start to charge an option-like bonus proportional to their returns in excess of the index benchmark, mathe-

[14] See, for instance, Jensen (1968); Sharpe (1991); Carhart (1997); French (2008); Fama and French (2010); Lewellen (2011); Busse et al. (2013).

matically, $u_a - u_b$. The above relationship does show that active fees are positively related to the excess performance of fundamental investing over factor investing.

4.3.2 Entry of skilled managers and informed households

The flexibility of the framework allows a mixed-strategy informed entry in an equilibrium. Taking price efficiency as given, I can figure out the equilibrium fraction of skilled managers, μ, as well as of informed households, ξ. I show that in an equilibrium, to guarantee fund flows for the overall active industry, there cannot be an excessive number of opportunistic managers.

First of all, the model shows that an endogenous amount of active managers incurs a cost k to become skilled. Condition (4.3.10) and previous results in Section 4.3.1 imply that

$$\mu = \frac{e^{\gamma k} - 1}{e^{\gamma k} - e^{-\gamma c}} \in \left[1 - e^{-\gamma k}, 1\right]. \tag{4.3.13}$$

One can see that μ decreases with respect to households' searching cost c. The harder it is for households to determine the managers' type, the more likely a manager opts to become opportunistic. In a hypothetical world where all households can easily find the skilled manager, all managers have to be truly skilled in order to guarantee fund inflows, i.e., $\mu \to 1$. However, in another extreme scenario, where it is impossible to determine the managers' type, not every manager free-rides. Instead, there is less, though not an insignificant amount of skilled managers, $\mu = 1 - e^{-\gamma k}$, as long as $k \neq 0$. This can be explained from the funds' perspective. First, the level of opportunism cannot be too high in order for active funds to secure inflows. Second, with most of the aggregation on the closet side, the flows left after apportion among competing funds amount to little; whereas on the skilled side it is less competitive.

Another implication is that the equilibrium fraction μ increases with the cost of fundamental information. When $k = 0$, $\mu = 0$; whereas when $k \to \infty$, $\mu \to 1$. This might seem surprising at first glance. But in this framework, the harder it is to obtain information, the more the managers are compensated by acquiring skills, implied by equation (4.3.11). This higher bonus facilitates a larger μ. At the same time, the high delegation cost incentivizes more households to become informed as they want to avoid selection error. Managers thus need to openly signal their superiority to secure flows from informed households. A detailed cost-benefit analysis will be provided in Section 4.5 of comparative statics.

Next I look into the equilibrium choice of asset managers. This helps us determine the equilibrium entry of informed households. If an active manager self-selects to become skilled, he attracts all the capital from informed households, and a fractional amount from

the uninformed counterparts, i.e., $\xi\,\lambda_a + \mu\,(1-\xi)\,\lambda_a$. These amount of inflows will be shared by a mass μ of skilled funds. The rest closet funds, with mass $(1-\mu)$, receive $(1-\mu)\,(1-\xi)\,\lambda_a$ amount of uninformed capital. In the equilibrium, coexistence of the fund managers implies that (i) an active fund is indifferent between being skilled and opportunistic, and (ii) a fund is indifferent between factor investing and fundamental investing strategies. These conditions yield the following mathematical equivalence:

$$(\alpha_a - k)\,\frac{\xi + \mu\,(1-\xi)}{\mu}\,\lambda_a = \alpha_a\,\frac{(1-\mu)\,(1-\xi)}{1-\mu}\,\lambda_a, \tag{4.3.14}$$

$$\alpha_b\,\lambda_b = \alpha_a\,\frac{(1-\mu)\,(1-\xi)}{1-\mu}\,\lambda_a. \tag{4.3.15}$$

Hence, condition (4.3.14), after simplification, yields the equilibrium fraction of households who learn the manager type through a costly search,

$$\xi = \frac{\gamma\,k\,\mu}{\gamma\,k\,\mu + \eta_I} \in [0,1]. \tag{4.3.16}$$

The equilibrium fraction of informed households (i) increases with managers' information cost k, (ii) decreases with households' searching cost c, (iii) is higher when the index price is more efficient, and (iv) is higher when there are more skilled managers. However, the value domain of ξ is different from that of μ. In the extreme case where price is completely inefficient, no household is willing to become informed. Whereas if price is perfectly efficient, or the managerial qualities are easy to figure out, all households become informed.[15] Price efficiency will be jointly determined by information costs, market turbulence, capital allocation decision, and information accuracy, with details in the next section.

4.3.3 Equilibria

I first derive the interior equilibrium in which factor investing, fundamental investing and direct investing coexist. But it could also be the case that households strictly prefer employing fund managers or even, one specific type of managers. The discussion of these boundary equilibria and boundary conditions follows afterwards.

The three indifference conditions of households (equation (4.3.8) to (4.3.10)), the fee condition (equation (4.3.11)), and two indifference conditions of managers (equation (4.3.14) and (4.3.15)), are sufficient to determine six equilibrium values. Given the capital allocation strategies of households, λ_h, λ_b and λ_a, I have figured out the equilibrium fees and

[15] In this extreme case when information of managers' type is costless, i.e., $c = 0$, all active managers have to become truly skilled ($\mu = 1$); And all households naturally become informed ($\xi = 1$). Thus condition (4.3.14) fails as there are no closet indexers. From $\xi = \mu = 1$ and other conditions, one can see $\lambda_b = \lambda_a$.

the amount of informed agents, $\{\alpha_a, \alpha_b, \mu, \xi\}$. Now I use the last two conditions, (4.3.8) and (4.3.15), to internalize the capital allocation. Then the individual and index price efficiency are endogenously determined.

Interior equilibrium

I term the most sophisticated equilibrium, where the three investment vehicles coexist, the *interior* equilibrium. Rearranging equation (4.3.15) implies that

$$\frac{\lambda_b}{\lambda_a} = \frac{\gamma\, k + \eta_I}{\gamma\, k\, \mu + \eta_I} > 1. \tag{4.3.17}$$

Since $\mu \leq 1$, regardless of the equilibrium allocation, demand for factor investing is higher, i.e., $\lambda_b \geq \lambda_a$. This is consistent with the latest tendency observed in the market, that, more capital flows into the index vehicles. The interior equilibrium is characterized by Proposition 4.3.1 below.

Proposition 4.3.1 (**Interior equilibrium**) *The interior equilibrium is characterized by the following capital allocation*

$$\lambda_b = \frac{\gamma\, \sigma_{q_2}}{\mu} \sqrt{\frac{1}{\left(\frac{1}{\sigma_e^2} + \frac{1}{\sigma_\varepsilon^2}\right)\left(e^{2\,\gamma\,(k+c)} - 1\right)}}, \qquad \lambda_a = \frac{\gamma\, k\, \mu + \eta_I}{\gamma\, k + \eta_I}\, \lambda_b; \tag{4.3.18}$$

the entry of skilled managers and informed households

$$\mu = \frac{e^{\gamma\, k} - 1}{e^{\gamma\, k} - e^{-\gamma\, c}}, \qquad \xi = \frac{\gamma\, k\, \mu}{\gamma\, k\, \mu + \eta_I};$$

and the delegation fees

$$\alpha_b = \frac{\eta_I}{\gamma}, \qquad \alpha_a = \frac{\eta_I}{\gamma} + k;$$

where η_I is uniquely given in the proof in Appendix Section D.3.

One direct observation is that higher demand for factor investing is usually accompanied by more closet indexers. Under this circumstance, selection mistakes based on prior information are more likely to be made when targeting active funds. The outcome can be explained by a time-consuming process of knowledge acquisition: active mutual funds have long been believed to have the managerial skills to generate excess returns. As the outcomes from past performance and academic research become widely accepted, people in turn become more convinced of the opportunism that exists within the active industry. The positive relation is, however, not always sustained, due to the endogeneity of closet indexing. The comparative statics will be analyzed in later sections.

Furthermore, I have shown in the previous section that, as long as k is not degenerate, in the equilibrium, there exists an intermediate lower bound for the number of skilled managers, mathematically, $\mu \geq 1 - e^{-\gamma k}$. The rationale holds for all the rest boundary equilibria as well. Finally, in the proof in Appendix Section D.3, I show the existence and uniqueness of the solution pair (λ_b, λ_a).

Delegation dominance

When exogenous variables take extreme values, the interior equilibrium, where three types of investment vehicles coexist, fails to hold. For example, when managers' cost of producing fundamental information drops, or the trades of individual stock become active ($\sigma_{q_2}^2$ rises), professional delegations are more attractive than self-investing due to overall market volatility and fee benefits. The equilibrium thus reduces to one that is taken at an intermediate boundary, where households completely resort to delegation instead of investing themselves, i.e., $\lambda_h = 0$ and $\lambda_b + \lambda_a = 1$.

In this case, condition (4.3.8) and (4.3.9) no longer bind yet households are still indifferent between factor and fundamental investing, i.e.,

$$u_a - c - \alpha_a = u_b - \alpha_b > u_h. \tag{4.3.19}$$

In total, I have six new conditions including (4.3.11), (4.3.10), (4.3.14), (4.3.15), (4.3.19), and $\lambda_b + \lambda_a = 1$, to solve for the six unknowns $\{\alpha_b, \alpha_a, \mu, \xi, \lambda_b, \lambda_a\}$. The analytical form of the delegation-dominating equilibrium is given below. Notice that the solutions of the fraction of skilled managers, μ, and of the capital in factor investing, λ_b, remain unchanged as in the interior equilibrium.

Proposition 4.3.2 (**Delegation dominate**) *The equilibrium where delegation dominates is characterized by the following capital allocation*

$$\lambda_b = \frac{\gamma\,\sigma_{q_2}}{\mu}\sqrt{\frac{1}{\left(\frac{1}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\xi^2}\right)\left(e^{2\,\gamma\,(k+c)} - 1\right)}}, \qquad \lambda_a = 1 - \lambda_b;$$

the entry of skilled managers and informed households

$$\mu = \frac{e^{\gamma\,k} - 1}{e^{\gamma\,k} - e^{-\gamma\,c}}, \qquad \xi = \frac{(2\,\lambda_b - 1)\,\mu}{(1 - \lambda_b)\,(1 - \mu)};$$

and the delegation fees

$$\alpha_b = \frac{k\,(1 - \lambda_b - \mu\,\lambda_b)}{2\,\lambda_b - 1}, \qquad \alpha_a = \frac{k\,(1 - \mu)\,\lambda_b}{2\,\lambda_b - 1}.$$

When fund delegation dominates direct investing, households play a mixed strategy between factor and fundamental investing. Here, factor investing is more popular, in that we still have $\lambda_b/\lambda_a = (\alpha_b + k)/(\alpha_b + \mu\, k) > 1$, similar to relation (4.3.17). The results show that popularity in delegated asset management drives up price efficiency. This is intuitive in that there is more informed trading in the market.

Moreover, with a rise in factor investing, for instance brought by larger supply uncertainty or lower fundamental information cost, the delegation fees, especially for active funds (α_a), decrease. This is consistent with the evidence found in Cremers et al. (2016), that active funds have to charge less in order to become the effective competitors against their factor investing rivals.

More importantly, by examining the exogenous changes, I figure out the necessary condition for an equilibrium where delegation completely dominates.

Lemma 4.3.2 (**Condition for delegation dominance**) *The necessary condition of an equilibrium where delegation dominates is*

$$\lambda_b = \frac{\gamma\, \sigma_{q_2}}{\mu} \sqrt{\frac{1}{\left(\frac{1}{\sigma_e^2} + \frac{1}{\sigma_\varepsilon^2}\right)\left(e^{2\,\gamma\,(k+c)} - 1\right)}} \in \Big(\frac{1}{2}, \frac{1}{\mu + 1}\Big], \tag{4.3.20}$$

where μ is given by (4.3.13).

Lemma 4.3.2 implies that when factor investing and fundamental investing coexist, there exists an upper limit for factor investing growth. This suggests that, despite the rising tendency of factor investing, it can never completely eliminate the other investment vehicles. The next section will present a clearer explanation of how the growth of factor investing is not sustainable.

Fundamental investing dominance

When fundamental information cost further reduces, or supply uncertainty or idiosyncratic risk further increases such that condition (4.3.20) no longer holds, the equilibrium where delegation dominates self-investing generates negative fees for factor investing, i.e., factor fund managers even provide interests to households, violating the feasibility of the equilibrium. Therefore, a new degenerate equilibrium arises, where households strictly prefer fundamental to factor investing, mathematically, $\lambda_b = 0$.

As is well known, before the emergence of indexing, people mainly counted on active equity funds or themselves to invest in risky assets. Thus an equilibrium where fundamental investing dominates factor investing is of interest with respect to the following questions:

(i) What gives rise to the entry of factor investing in a world containing only fundamental investing and direct investing? (ii) Can factor investing grow immoderately to seize the entire market?

Let us first characterize the equilibrium with $\lambda_b = 0$. Because factor funds have no inflows, it becomes non-profitable for deploying factor strategies. Hence the coexistence of two types of funds, condition (4.3.15), fails. Coexistence of direct investment and active delegation implies that

$$u_a - c - \alpha_a = u_h > u_b - \alpha_b.$$

With the rest conditions, I character the equilibrium where fundamental investing dominates factor investing in the following proposition.

Proposition 4.3.3 (**Fundamental dominate**) *The equilibrium where fundamental investing dominates is characterized by the following capital allocation*

$$\lambda_a = \min\left[\gamma\,\sigma_{q_2}\sqrt{\frac{1}{\left(\frac{1}{\sigma_e^2}+\frac{1}{\sigma_\varepsilon^2}\right)\left(e^{2\,\gamma\,(k+c)}-1\right)}}, 1\right], \qquad \lambda_b = 0;$$

the entry of skilled managers and informed households

$$\mu = \frac{e^{\gamma\,k}-1}{e^{\gamma\,k}-e^{-\gamma\,c}}, \qquad \xi = 1;$$

and the delegation fees

$$\alpha_b = 0, \qquad \alpha_a = k.$$

From Proposition 4.3.3, one can tell that the price of the market portfolio has become perfectly efficient. Thus there is no difference between investing directly and employing factor funds. However, active funds still provide price discovery and hedging functions to households. Nevertheless, active funds cannot earn much by acquiring fundamental information when k is small, they lack the incentive to be truly skilled thus tend to free-ride. All households thus perform enough due diligence to ascertain the manager quality due to the high level of opportunism in the industry, and allocate money to the skilled managers. Upon the realization of some extreme conditions (for instance, where information production for active managers is too easy or the supply of individual assets is too volatile), households could even play a purely-active strategy in the equilibrium ($\lambda_a = 1$).

The condition to sustain the fundamental-dominating equilibrium is given in Lemma 4.3.3. The switching point of the two equilibria, may, at a first glance, be discontinuous. However, I show in the proof of the following lemma that the equilibrium solutions are always *upper-*

semicontinuous. In other words, multiple equilibria coexist along the boundary line of $\lambda_b = 1/(\mu+1)$ and $\lambda_a = \mu/(\mu+1)$.

Lemma 4.3.3 (**Condition for fundamental dominance**) *The necessary condition of an equilibrium where fundamental investing dominates is*

$$\lambda_a = \gamma\,\sigma_{q_2}\sqrt{\frac{1}{\left(\frac{1}{\sigma_e^2}+\frac{1}{\sigma_\varepsilon^2}\right)\left(e^{2\gamma(k+c)}-1\right)}} \in \left(\frac{1}{\mu+1},1\right], \tag{4.3.21}$$

where μ is given by expression (4.3.13).

A natural concern arises that whether there exists an equilibrium where factor investing dominates the rest. In this framework, because all the agents endogenously select their roles, such equilibrium does not exist. But relaxing this requirement, the insight of this question is in essence identical to the situation modeled in Gârleanu and Pedersen (2018). With nobody employing active managers, the problem reduces to one with only two kinds of players—informed (corresponding to factor investors in my model) and uninformed (corresponding to direct investors in my model). The degenerate equilibrium is not the main focus of this paper.

4.3.4 Ceiling of factor investing

Following the above analysis, there indeed exists an implicit "ceiling" for factor investing: the fraction of capital in factor investing cannot exceed a fraction of $1/(1+\mu)$ of the whole market share, as implied by the estimated conditions of the delegation-dominating equilibrium and the fundamental-dominating equilibrium. The simplistic relation between λ_b and μ suggests a feasible measure, irrespective of the composition of risk appetite and information costs, to central planners concerning how much capital they shall expect to be in factor investment vehicle. Bebchuk and Hirst (2019) raise the concern that index funds control an increasingly large amount of voting shares in S&P500 companies. Regarding this as a crucial concern, the authors estimate how big can index funds grow and discuss its effect on ownership patterns in public markets from a corporate governance perspective.

In my model, the ceiling of factor investing is largely determined by the opportunism level in the active industry. For active funds, being closet is a bargaining: if all households are uninformed, it is worthy; if, on the contrary, households all choose to learn manager type, then funds have to become truly skilled to retain capital. Therefore, the amount of closet funds depends both on households' searching cost and managers' information cost. So does the ceiling: the upper bound of factor investing is lower when it is harder for

managers to obtain fundamental information, or when it is easier for households to detect closet indexing.

The result yields certain policy evaluations. If, regulators engage in improving market efficiency by finding out and revealing closet indexing to the public, then all active managers have to be truly skilled, i.e., $\mu = 1$. In this case factor investing could at most occupy half of the market shares. If, on the contrary, opportunistic managers exist extensively, the ceiling goes up. However, it can never reach 100%. In order for active managers to sustain inflows, there cannot be an excessive number of closet indexers, precisely, $\mu \geq 1 - e^{-\gamma k}$.

In the equilibria, households' capital allocations are determined by a bunch of exogenous variables, $\{k, c, \gamma, \sigma_{q2}^2, \sigma_e^2, \sigma_\varepsilon^2\}$. Hence, the upper bound, $1/(\mu+1)$, is an endogenous moving surface. This can be seen from Figure D.3 in Appendix Section D.1, which summarizes the shift between equilibria when the exogenous variables are fixed except for information costs (k, c). Furthermore, there exist both situations where factor investing and fundamental investing move in (i) the same and (ii) the opposite direction. Hence, this model not only addresses the univariate effect of ingredients, but also explains their joint influence on the stylized facts of equilibrium capital allocation.

The model proves that multiple equilibria coexist along the boundary line of $\lambda_b = 1/(\mu+1)$ and $\lambda_a = \mu/(\mu+1)$. In fact, any combination of (λ_h, λ_b) satisfying $\lambda_b + \lambda_h = 1/(\mu+1)$ is a feasible equilibrium. Therefore, the shift between three kinds of equilibria is always smooth.

The reason behind the results is of a model interaction. The more people investing in factor funds, the more market-level information is impounded into asset prices. Upon rising to the peak level, factor investing has fully exploited the informational advantage over the systematic factor. Hence households can no longer benefit by putting money into factor funds. However, fundamental investing still has informational advantage over idiosyncratic loadings because of the great anomalies in individual security. This is expected to increase the relative gains from acquiring fundamental information, which leads to an endogenous shift from factor investing to fundamental investing. Hence, in the model, there comes a limit of how big factor investing can grow.

4.4 Household Heterogeneity

A fundamental insight into the asset management industry is to understand the true nature of the demand side. Heterogeneity thus matters because it generates a market where investors could have different demand schedules. To see if the allocation strategy and other equilibrium characteristics persist in a more realistic setting, in this section, I

link the standard REE model to household finance. The model can help understand the effect of *household heterogeneity* on equilibrium capital allocation and the structure of the delegation industry. I show that household heterogeneity provides a possible explanation for the rise in factor investing and the compression of fund fees, a link that is testable empirically.

In the investment market, households can differ in their initial wealth and in their cognitive ability in relation to the detection of managerial skills. I hereby derive the equilibrium relaxing the identical household assumption in the baseline model. I will distinguish in my setup between *sophisticated* and *unsophisticated* households, and discuss separately their capital allocation strategies. Heterogeneity implies two main differences: (i) each individual may have specific preferences towards asset management vehicles, though in aggregation the three types of investing coexist in the market; and (ii) with heterogeneous wealth level, each individual's contribution to the funds' assets under management can differ.

I assume a continuum of households, labeled by $n \in \mathbb{N}^+$. Consider the statistical structure of different households, characterized by the set of their information cost, risk appetite and initial wealth, $\{c^n, \gamma^n, W^n\}$. Similar to the assumption in Peress (2003) that absolute risk aversion decreases with wealth, in this framework, the constant *absolute* risk aversion in exponential utility, γ^n, is determined by a following relation with *relative* risk aversion γ^R: $\gamma^n = \gamma^R / W^n$.[16] Because the difference in relative risk aversion is negligible compared to the variation in wealth, I simply assume all the households share an identical γ^R. Thus one can regard the difference in absolute risk appetite directly caused by the variation in initial wealth of households, with wealthier households less absolutely risk-averse, and vice versa. Therefore, each household can be uniquely characterized by a set $\{c^n, W^n\}$. The information cost and wealth distributions satisfy $\mathbb{E}[c^n] = c$ and $\mathbb{E}[W^n] = W$.

Following the wealth and cost distribution, the fees charged and the payoffs delivered by the funds also differ from one household to the next. Typically, wealth affects households' certainty-equivalent utility generated by the fund managers, $\{u_a^n, u_b^n\}$, because of its effect on risk appetite. Despite the above asymmetry, all households share a common price of assets, price efficiency, and funds strategies ($\{\boldsymbol{x}_a, \boldsymbol{x}_b\}$), determined by the market equilibrium. Denote by $\mathbb{E}[\frac{1}{\gamma^n}] = \frac{W}{\gamma^R} = \frac{1}{\gamma}$.

[16]The link between wealth and risk aversion mimics the property of the more standard CRRA utility, while maintaining the tractability of a CARA-utility model. This modeling strategy is similar to Saha (1993), Makarov and Schornick (2010), Koijen et al. (2019), and Pedersen et al. (2019). The relation is also documented empirically by Gomez et al. (2019).

4.4.1 Who favors factor investing?

In this section, I defer all the detailed derivation to Appendix Section D.4, and provide corresponding outlines and results. Fundamental investing is a better reply to direct investing for a household if and only if

$$u_a^n - c^n - \alpha_a^n \geq u_h^n.^{17}$$

Thus the fraction of informed households, ξ, is determined endogenously by

$$\xi = \mathbb{P}(u_a^n - c^n - \alpha_a^n \geq u_h^n) = \mathbb{E}[\mathbb{1}_{\{u_a^n - c^n - \alpha_a^n \geq u_h^n\}}]. \tag{4.4.22}$$

Note that with a distribution of fixed cost or wealth, the market effectively has an upward-sloping supply curve, which generates smooth entry of informed households. Assume that each household has to pay an endogenous delegation fee proportional to her submitted capital. Then *ceteris paribus*, applying similar analysis as in Section 4.3.1 yields

$$\alpha_b^n = \frac{\eta_I}{\gamma^R} W^n, \qquad \alpha_a^n = \Big(\frac{\eta_I}{\gamma^R} + k\Big) W^n.$$

Inheriting the same notation for the fraction of skilled managers, μ, the model claims that the rest of the households cannot gain from selecting an active manager based on prior information, or from investing through factor funds, mathematically,

$$\mu\, e^{-\gamma^n (u_a^n - \alpha_a^n)} + (1-\mu)\, e^{-\gamma^n (u_b^n - \alpha_a^n)} = e^{-\gamma^n (u_b^n - \alpha_b^n)} = e^{-\gamma^n u_h^n}.$$

Hence, individuals with heterogeneous wealth and cognitive ability favor different investment vehicles, with their preferences and classifications specified in the following lemma.

Lemma 4.4.1 (**Who favors factor investing**) *A household is classified as sophisticated if, her statistical structure* $\{W^n, c^n\}$*, satisfies*

$$\frac{c^n}{W^n} \leq -\frac{1}{\gamma^R} \log\Big(\frac{1 - e^{\gamma^R k}(1-\mu)}{\mu}\Big). \tag{4.4.23}$$

[17] The strategy within a single type of funds is identical, yet yields different certainty-equivalent wealth to individuals. Utility u_j^n $(j = h, b, a)$ to household n is determined by:

$$e^{-\gamma^n u_j^n} = \mathbb{E}\left[\exp\Big\{-\frac{1}{2}\left(\mathbb{E}[\boldsymbol{v}|\mathcal{I}_j] - \boldsymbol{p}\right)^\intercal \boldsymbol{\Sigma}_j^{-1}\left(\mathbb{E}[\boldsymbol{v}|\mathcal{I}_j] - \boldsymbol{p}\right)\Big\}\right],$$

following definition (4.2.2).

(i) The sophisticated households (ones with high wealth level and cognitive ability such that equation (4.4.23) *satisfies) strictly prefer fundamental investing;*

(ii) The unsophisticated households (ones with low wealth level and cognitive ability such that equation (4.4.23) *violates) stay uninformed and are equally likely to choose direct investing, factor investing, and allocating to active funds based on prior information.*

The outcomes of the likely composition of fundamental investing households are straightforward and intuitive, in that households with a preference for active delegation should be relatively more sophisticated with either large wealth endowment or special insights into the fund manager capabilities. In reality, for instance, institutions or rich individuals rely heavily on active mutual funds, or even on alternative investment including private equity and hedge funds. They largely exert resources to examine the managers through assessing their investment process, trading infrastructure, etc. However, small retail investors are more conservative and thus value direct investing and factor investing more. High management fees and uncertainty about managerial quality create a barrier for them to enter the active industry. As claimed by Dyck et al. (2013), it is indeed the case that sophisticated institutional investors are more likely to use active asset management compared to retail investors, as the latter group faces much higher investment costs.

Note that in the baseline model with identical households, demand for factor investing is higher than for fundamental investing, as long as they coexist. With heterogeneous households, because (i) they have different preferences towards funds, and (ii) those that choose active asset management are less wealth constrained, it could be the case that the total amount of money under management of active funds is larger than that under factor funds.

4.4.2 Effects on capital allocation

In contrast to the homogeneous baseline model, though factor investing, fundamental investing and direct investing coexist, different types of funds attract different kinds of households, whose wealth affects the ultimate payment to the managers. Skilled managers receive all of the inflows from the sophisticated households, and part of the inflows from the unsophisticated households; whereas factor and closet funds, as given in Lemma 4.4.1, only attract the uninformed and unsophisticated households. Hence the equilibrium conditions

of fund managers become

$$\mathbb{E}[\alpha_a^n W^n|\text{uninformed}]\,\frac{(1-\mu)\,(1-\xi)}{1-\mu}\,\lambda_a = \mathbb{E}[\alpha_b^n W^n|\text{uninformed}]\,\lambda_b = \left(\mathbb{E}[\alpha_a^n - k\,W^n|\text{informed}]\,\frac{\xi}{\mu} + \mathbb{E}[\alpha_a^n - k\,W^n|\text{uninformed}]\,\frac{\mu\,(1-\xi)}{\mu}\right)\lambda_a.$$

With these indifference conditions, one can solve the equilibrium system. λ_a and μ are both functions of price inefficiency η_I. ξ can be uniquely mapped to μ, thus is also a function of η_I. Hence I determine the equilibrium price efficiency first, then all the rest terms follow. The detailed derivation is deferred to Appendix Section D.4.

I now examine how household heterogeneity affects the equilibrium capital allocation, fees, and opportunism level in the economy. As we are faced with a joint distribution of two variables, wealth level and cognitive ability (information cost), one needs to specify the distribution to model the effects.

Wealth inequality

First, fix the cognitive ability dimension and consider wealth inequality. Assume a fixed information cost for every household, $c^n \equiv c$. To model the dispersion of wealth, I restrict the distribution of wealth, W^n, on a fixed support, $[0, 2\,W]$. Now I fix the expectation of the distribution to be W, and vary the variance and shape of distribution to capture the dispersion. Without loss of generality, I assume the wealth variable W^n satisfies the Beta distribution on the finite positive support $[0, 2\,W]$, with density function as follows,

$$\text{Beta}(a,b) = \frac{(\frac{x}{2\,W})^{a-1}\,(1-\frac{x}{2\,W})^{b-1}}{2\,W\,B(a,b)}, \qquad \text{where } B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a,b)},$$

where B stands for the Beta function and Γ stands for the Gamma function. The mean of this distribution is $2\,W\,a/(a+b)$. To focus on the dispersion, I fix the mean at W by assuming $a = b$. When $a = b < 1$, the distribution is bimodal (or "U-shaped") thus is termed highly dispersed; whereas when $a = b > 1$, the distribution is unimodal, with a higher a implying a more concentrated distribution. In the intermediate scenario where $a = b = 1$, the distribution becomes uniform. In a nutshell, the smaller a is, the more unequal the wealth distribution is. Furthermore, Gini coefficient is a well-known measure for wealth inequality, which, given any wealth distribution, can be calculated theoretically.[18] Thus more intuitively, I map the wealth distribution to Gini index, and

[18]Gini coefficient is a value taken within the region $[0, 1]$. A coefficient of 1 is interpreted as maximal inequality and a coefficient of 0 means households have completely identical wealth. Gini index has long been used to measure wealth inequality, see, for instance, Yitzhaki (1979) and Lambert and Aronson

use it as the main measure of statistical wealth dispersion. See Figure D.4 in Appendix Section D.4.3 for an intuitive summary.

One can tell the following equilibrium features brought by wealth inequality.

Lemma 4.4.2 (**Wealth inequality**) *As the wealth level among households becomes more dispersed:*

(i) More managers become skilled;
(ii) Fewer households become informed towards manager quality;
(iii) The percentage delegation fees decrease for both factor and active funds;
(iv) The demand for factor investing increases;
(v) Prices are more informative.

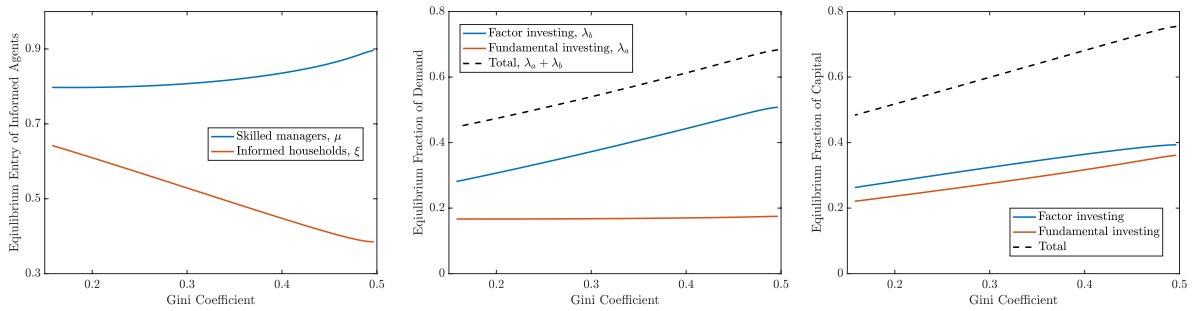
Somewhat surprisingly, when the population in a society has a larger wealth gap, the demand for factor investing is higher. By unraveling the rationale behind this, one begins to detect some logic. First, wealthier households benefit from searching for fundamental investors, since their searching cost is low relative to their capital. Second, heterogeneity concentrates wealth in the hands of less marginally risk-averse wealthy households, whose wealth effect brings the individual prices closer to their fundamental values, shrinking the outperformance of fundamental investing. At the same time, the large possible inflow from sophisticated households makes it more attractive for managers to become skilled. Thus many active managers rush to acquire relevant skills, thereby increasing the inside competition of fundamental investing and curtailing the return of active strategy even further. The equilibrium outcome of decreasing returns to scale provides justification for such assumption in Berk and Green (2004) and Pástor and Stambaugh (2012).

Therefore, we observe simultaneously that: (i) active funds lower their fees to remain competitive; and (ii) people become less interested towards fundamental investing. The aggregation effect is, at first glance, ambiguous. But the model shows that the latter effect dominates: more households thus, instead, choose to stay uninformed and embrace the cost-efficient factor investment vehicle, represented by a smaller ξ. Furthermore, the reduced fees earned by active funds force factor funds to lower their fees as well, leaving less wealthy households less likely to directly invest and more likely to opt for delegation.

Active funds, on the one hand, are directly harmed by the decrease of informed households; on the other hand, they can attract more of the random flows from the uninformed households because of the fee advantage. Hence, in aggregation, the demand for fundamental investing remains roughly unaffected, suffering only a tiny decrease. The rise in demand for factor investing is mainly compensated by direct investing.

(1993).

On the contrary, when most households are middle class, factor investing diminishes, and converges to the limit of λ_b in the homogeneous-wealth equilibrium (if $W^n \equiv W$ for any n), given by equation (4.3.18). For clarity, a numerical example is given in Figure 4.1. With a reasonable choice of parameters, in the example the delegation fees decrease by about 0.5% for both funds as the wealth distribution turns from equal to unequal.[19]



A. Informed agents B. Demand C. Capital

Figure 4.1: *Wealth inequality on capital allocation.* I vary the parameter of Beta distribution, a, in $(0, 6]$ to get the corresponding Gini coefficient (x-axis), holding the expected wealth $W = 8$ and the support of wealth distribution $[0, 16]$. For the other parameters, take $\gamma^R = 3$, $k = 0.05$, $c = 0.1$, $\sigma_{q_I}^2 = \sigma_{q_2}^2 = \sigma_e^2 = \sigma_s^2 = 2\,\sigma_\varepsilon^2 = 0.5$.

Panel B of Figure 4.1 confirms the previous analysis of increasing demand for factor investing. Nevertheless, due to wealth effects, the money put into funds by rich and poor households carry different levels of power. Thus the total capital, in monetary terms, pouring into funds differ from one household to the next. Skilled active funds have a wealthier investor base. The model shows that the outcome of the rise in factor investing is robust in terms of capital, and that heterogeneity amplifies the fraction of capital in both funds (Panel C). In other words, wealth inequality among households increases the assignment of professional delegation while curtailing uninformed direct investment.

In the past decades, tendencies of increased factor investing, decreased delegation fees and increased wealth inequality have all been observed. Typically, for instance, the U.S. investment market has been experiencing these changes simultaneously, as can be seen from Figure D.1 and D.2. This model to some extent ties the three observations together and explores the possible causal link between these trends over time.

Heterogeneous cognitive ability

Then I look into the other dimension—heterogeneous cognitive ability among households, following the same rationale. By high cognitive ability, I mean the household can incur

[19]The numerical level of relative risk aversion is consistent with that in Gârleanu and Pedersen (2019).

a low cost to become aware of the manager quality. This time I fix the wealth level, $W^n \equiv W$, and model the cost dispersion on a fixed, finite, and positive support $[0, 2\,c]$. Without loss of generality, I assume $c^n \overset{d}{\sim} \text{Beta}(a, b)$ with $a = b$. Thus the expected cost of learning the managers' type is fixed at c. A small a stands for asymmetric insight of manager skills. In real world, it is highly possible that institutions and professional investors have more insight towards which active manager is better.

One can tell the following equilibrium features with households heterogeneity in cognitive abilities, as given in Lemma 4.4.3.

Lemma 4.4.3 (**Cognitive ability dispersion**) *As the cognitive ability among households becomes more dispersed:*

(i) Fewer households become informed;
(ii) Fewer managers become skilled;
(iii) Demand/Capital switches a little bit from fundamental investing to factor investing.

A numerical example is given in Appendix Section D.4.3. In Grinblatt et al. (2011), the authors find that high-IQ households are more likely to delegate to funds and hold a large number of stocks. In this model, I enrich the analysis by looking into the discrepancy in cognitive ability and its effect on delegation in the intensive margin. The model shows that a larger inequality in cognitive ability gives rise to factor investing. This tendency can be explained by a higher level of opportunism among fundamental investors, which prevents households from being informed and delegating money to such investors, as represented by a smaller ξ. However, the implication is not as clear here as with wealth inequality, and the effect is not as large. Furthermore, an asymmetric cognitive ability across households amplifies the average outperformance of fundamental investors in equilibrium. This is because a decreasing amount of skilled active managers lowers the inside competition in fundamental investment.

By examining a series of reasonable exogenous variable choices, I conclude that, in this framework, the wealth effect dominates the cognitive ability effect in terms of magnitude. However, there exist other possibilities in relation to modeling cognitive ability. For instance, it is highly likely that households with different cognitive capability have different certainty-equivalent wealth generated by funds. The possibility that cognitive ability affects utility is not incorporated in my model. Had this been included, it is possible that other effects, maybe more magnificent, on the overall asset market equilibria brought by cognitive capability could have emerged.

Unequal society

I analyze independently the two dimensions of heterogeneity in the previous sections. Now I examine the case where households' wealth and cognitive capability are correlated. The link between the two, in the existing literature thus far, is a bit mixed. Rindermann and Thompson (2011) find some extent of positive correlation, whereas Zagorsky (2007) find a quadratic relationship. In this framework, it is highly likely that one with higher wealth endowment also has higher cognitive ability (or, lower information cost). For instance, if one thinks of "households" in a broader sense to include institutions, pension funds, endowments and sovereign wealth institutes, these asset owners aggregate large amount of wealth, and are highly likely to have more insight into the fund quality. I term such a society with positive correlation between wealth and cognitive ability an *unequal* society. Equivalently, it can be captured by a negative correlation between wealth and information cost level.

For this purpose I take one of the Archimedean copula—Ali-Mikhail-Haq copula (AMH for short, see Ali et al. (1978))—as the joint distribution of (W^n, c^n), which could capture both positive and negative correlations. Restricting the marginal distribution of wealth as uniform on $[0, 2\,W]$, and the marginal distribution of information cost as uniform on $[0, 2\,c]$, the AMH copula implies the following cumulative distribution function:

$$\mathrm{AMH}(\theta) = \mathbb{P}(W^n \leq x, c^n \leq y) = \frac{x\,y}{4\,W\,c - \theta\,(2\,W - x)\,(2\,c - y)}, \qquad \text{where } \theta \in [-1, 1].$$

The parameter θ captures the correlation between wealth and information cost. A negative θ implies a negative correlation, intuitively, wealthier households have easier access to the managerial quality information; A positive θ implies a positive correlation; And $\theta = 0$ means that wealth and cognitive ability are independently uniformly distributed. Note that, however, $\theta = 1$ (-1) is not translated to perfectly positive (negative) correlation, but a one to one mapping to some intermediate correlation level.

The asset market equilibrium is influenced by the correlation in the following way.

Lemma 4.4.4 (**Unequal society**) *As the society becomes more unequal:*

(i) More managers become skilled;
(ii) Fewer households become informed towards manager quality;
(iii) The percentage delegation fees decrease for both factor and active funds;
(iv) The demand for factor investing increases;
(v) Prices are more informative.

The intuition and outcomes are very much similar to the case with one-dimensional dispersion in wealth. A numerical example summarizes the findings, as provided by Figure D.6 in Appendix Section D.4.3.

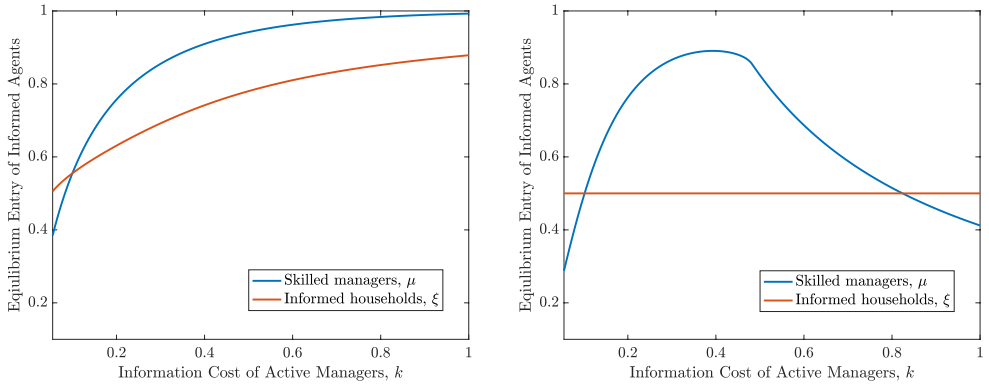
4.4.3 Dynamics of opportunism

Heterogeneity, or an unequal society, can help explain other pertinent features of the asset market equilibria, including the dynamics of opportunistic managers.

Consider a discrete example below, with only two different types of households. The first group of households are relatively rich (with wealth $\overline{W}$) and all become informed due to their easy accessibility to inside information of manager type (for instance, they share a searching cost normalized to $\underline{c} = 0$). The second group owns limited wealth ($\underline{W}$) yet an extremely high searching cost ($\overline{c}$), thus are never able to learn the type of active managers. The market consists of mass ξ of the first group and mass $1 - \xi$ of the second group. Lining up with the previous terminologies, the first group coins the sophisticated households and the second group unsophisticated. The Bernoulli distribution exogenously determines the entry of informed households, ξ. Any exogenous change in the investment market, including frictions or supplies, cannot affect the amount of informed households. A skilled manager cannot attract more households through openly signaling his type, either. Typically, in the following example, I show that a variation in the information cost of fund managers, k, though not affecting household partition, has interesting implications on the dynamics of opportunism in the active industry.

Example 4.4.1 *As shown in Lemma 4.4.1, the group of sophisticated households always prefer fundamental investing, as they can target the "right" managers. While the unsophisticated group who lacks both money and insight, randomizes between directly investing, factor investing or picking active funds based on prior information. If they choose active delegation, there is a certain probability* $1 - \mu$ *that their money goes to an opportunistic manager. Holding the average wealth and information cost consistent with the homogeneous model, I compare the change of informed agents with respect to managers' information cost* k*, as shown in Figure 4.2.*

Interestingly, in such an extremely unequal society, the most efficient level of active industry—where largest amount of managers are truly skilled—becomes intermediate as a function of k*. I term the market with heterogeneous households a "talent market," where a fixed fraction of sophisticated households knows clearly the managers' type, yet the rest are prevented from accessing such information due to their high searching cost. The amount of talented households,* ξ*, is exogenous. The homogeneous equilibrium, however, happens in*



A. Homogeneous B. Heterogeneous

Figure 4.2: *Comparison between homogeneous and heterogeneous market.* As I vary the choice of k in $(0, 1]$, this figure shows the change of equilibrium amount of skilled managers and informed households, with Panel A for homogeneous equilibrium and Panel B for heterogeneous equilibrium. I keep mean wealth and cognitive capability the same in two models: $\mathbb{E}[c^n] = \frac{1}{4}$ and $\mathbb{E}[W^n] = 2$. In Panel B I assume $\overline{c} = 0.5$, $\underline{c} = 0$, $\overline{W} = 3$, $\underline{W} = 1$. For the other parameters, take $\xi = 0.5$, $\gamma^R = 4$, $\sigma^2_{q_I} = \sigma^2_{q_2} = \sigma^2_e = \sigma^2_s = 2\,\sigma^2_\varepsilon = 0.2$.

an "effort market," where every household is identical and can exert some effort to search for managers, in which case, endogenously determines ξ.

In the effort market, as the information cost of managers, k, increases, the managers also charge more to compensate the cost. Thus households who choose active delegation need to be more confident in picking the right managers. Hence, the entry of informed households is monotonically increasing with respect to k. The large amount of informed households also drives up the amount of skilled managers, because they want to openly signal their type in order to attract money from the informed households. The results are illustrated in Panel A of Figure 4.2.

In contrast, in the talent market, the monotonicity no longer exists, as shown in Panel B. When information cost k is relatively small for the active managers, they do benefit from acquiring skills and charging a premium accordingly. Nonetheless, at some point they start to give up the inflows from the talented households, in the model, μ shrinks. This is because the new inflows, after being shared by a large fraction of competing skilled funds, cannot offset the high cost of acquiring skills. They also need to generate large enough gross returns to justify the high fees. Because of the exogeneity of ξ, they can never incentivize the unsophisticated households to become informed. This is intrinsically why acquiring skills ceases to be attractive for managers in the heterogeneous world. As a consequence, μ is hump-shaped.

The outcome above, that the entry of skilled managers differs in the *effort* and *talent* market, is intuitive. In the effort world, μ is endogenous but uniquely determined by the

information frictions:

$$\mu = \frac{e^{\gamma k} - 1}{e^{\gamma k} - e^{-\gamma c}}.$$

The variation in k generates a monotonic effect on μ. A higher fundamental information production cost allows active managers to charge higher premium and to benefit from information acquisition. Hence being a fundamental investor becomes more attractive.

In contrast, in the talent world, μ is a function of both information frictions and price efficiency. For instance, in the above extreme example with Bernoulli-distributed statistical structure of households,

$$\mu = \frac{\eta_I}{k\,\gamma^R}\,\frac{\overline{W}\,\xi}{\underline{W}\,(1-\xi)},$$

where k influences μ through (i) influencing price efficiency, and (ii) influencing μ directly. When k increases from a low level, there is high demand for factor investing, forcing η_I to be low. The highly informative index price provides space for fundamental investing to profit. Thus managers are incentivized to become skilled. However, a too large k prevents households from using funds and gives birth to opportunistic managers who try to attract part of the capital without incurring the high cost. Thus there exists complex involution between k and μ, causing the monotonicity to vanish.

In fact, in a continuous talent market, where wealth and information cost satisfy a continuous joint distribution, the rationale of analyzing μ is similar as in the discrete example. The entry of skilled managers will be determined through finding the fixed point of the following equation:

$$\frac{\eta_I}{\gamma^R\,\mu}\,\mathbb{E}[W^n|\text{informed}]\,\xi = k\,\mathbb{E}[W^n|\text{uninformed}]\,(1-\xi),$$

where ξ is no longer exogenous yet is uniquely determined by μ (details in Appendix Section D.4). Thus given η_I, the above expression uniquely determines μ, in a similar fashion as in the discrete example. In a nutshell, household heterogeneity makes the structure of the active industry more intertwined with the market conditions, thereby alters the dynamics of equilibrium capital allocation and delegation fees.

4.4.4 "Alpha"

Household heterogeneity has certain implications on the outperformance of fundamental investing over factor investing. One canonical way to estimate a skilled active fund's excess return over market is through the Capital Asset Pricing Model (CAPM) and look into the "alpha" it generates, i.e., the intercept of the Security Market Line (SML). In this model, the only group of participants who can capture the abnormal return is the group of skilled

active managers, who yield inside information towards idiosyncratic loadings and profit through rigorous stock selection. They are also known for charging higher fees due to the alpha they generate.

Based on the model setup, the risk-appetite-adjusted outperformance of skilled active funds over factor funds, $\gamma^n (u_a^n - u_b^n)$, becomes a tractable measure that captures the excess performance in a stylized fashion.[20] From the above analysis, the abnormal return of fundamental over factor investing can be linked to the equilibrium amount of skilled managers as

$$\gamma^n (u_a^n - u_b^n) = \gamma^R k - \log\Big(\frac{1 - e^{\gamma^R k}(1-\mu)}{\mu}\Big).$$

Because the relative risk aversion and managers' information cost are exogenous, the outperformance level varies in the opposite direction of μ, another endogenous variable. As μ increases, the more efficient prices and the higher competition among skilled managers lead to an endogenous decreasing returns to scale of fundamental investors. Therefore, the highest possible excess return generated by fundamental investing is when there is least competition within the industry.

4.5 Drivers and Effects of Rising Factor Investing

The analysis of asset market equilibrium and household heterogeneity provides one possible explanation for the rise in factor investing. In this section, I summarize some other possible drivers of factor investing through analyzing the comparative statics of the equilibria. These can help understand the empirical observations as well. Given the analytical equilibrium solution of demand for factor investing in Proposition 4.3.1, I present a preview of the contributes in the following proposition, and discuss the intuition in the following sections.

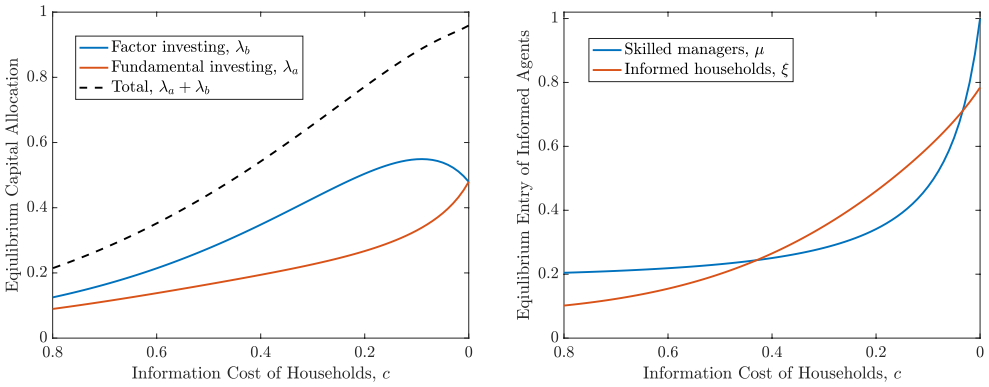
Proposition 4.5.1 (**Drivers of factor investing**) *In the interior equilibrium, demand for factor investing:*

(i) increases then decreases with household searching cost c;
(ii) decreases with managers' cost of fundamental information k;
(iii) increases then decreases with investor risk aversion γ;
(iv) increases with idiosyncratic risk σ_e^2 and supply uncertainty $\sigma_{q_2}^2$;
(v) increases with household inequality.

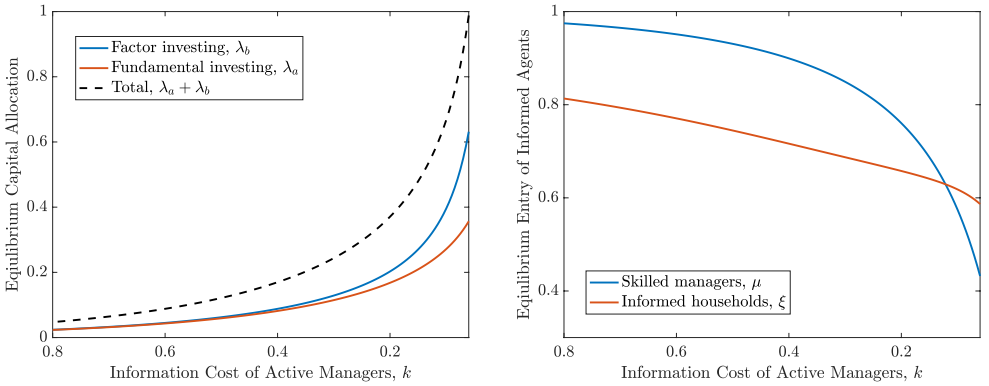
[20] In Appendix Section D.5, I provide the rationale to properly analyze "alpha" in CAPM.

Information costs

In general, a drop in information costs, either c or k, increases the demand for fund delegation, which is compensated by a decrease in direct investment. However, despite this general tendency, the detail in and rationale behind how the two types of information cost influence capital allocation are very different. Specifically, the two types have opposite effects on the level of opportunism in the economy. The comparative statics can be found in Figure 4.3.



A. Households' information cost, c



B. Managers' information cost, k

Figure 4.3: *Comparative statics w.r.t. households' and managers' information cost.* As I vary $c \in [0, 0.8]$ while fixing $k = 0.07$ (Panel A), or vary $k \in (0, 0.8]$ while fixing $c = 0.1$ (Panel B), this figure shows the equilibrium capital allocation, and amount of skilled managers and informed households. For the other parameters, take $\gamma = 3$, $\sigma^2_{q_I} = \sigma^2_{q_2} = \sigma^2_e = \sigma^2_s = 2\,\sigma^2_\varepsilon = 0.2$.

First, let us take a look at the searching cost of finding out the managers' type. The demand for factor investing is hump-shaped with respect to the change of c. A general decrease in searching cost c has the following two effects: (i) it makes opportunistic strategy harder, curtailing the amount of closet indexers in the economy; and (ii) it shrinks the outperformance of fundamental investing compared to factor investing (from equation (4.3.8) and (4.3.9) one can conclude that $u_a - u_b = k + c$). In turn, fund managers

have to charge less, leading to compression of both fund fees. Therefore, households are more likely to delegate their money to funds in general, especially the cost-efficient factor investing vehicle. Ultimately, more informed trading improves price efficiency, both for the index and individual assets. However, when searching cost c further decreases, this regime may backfire. A sufficiently small cost c leads to a sharp increase in the amount of skilled managers, leading households to shift to fundamental investing as a "substitute" due to the high efficiency in the active industry. In the extreme case that households can ascertain managerial quality without any cost ($c \to 0$), all active managers have to acquire skills in order to attract inflows ($\mu \to 1$). Thus the efficiency of the active industry drives its inflows to the maximum level, which is equal to the flows of factor funds, i.e., $\lambda_b = \lambda_a$.

Second, for comparison, I analyze the effect of managers' cost of obtaining fundamental information, k. With a large k, fundamental investors struggle to produce inside information. Thus, they charge more and these high expenses block most households from engaging professional delegation. Therefore, in the feasible region of the interior equilibrium, a natural outcome is that the decrease in managers' information cost spurs more fund delegation. When k is relatively small, active funds do not gain much for becoming skilled. This is because the highly efficient asset prices shrink the profitable region of fundamental investors, making them reluctant to become skilled. Due to the high level of opportunism and the tiny gap between the fees for fundamental investing compared to factor investing, there is a not insignificant number of households willing to find and vet the skilled managers and delegate money to them. Note that I do not emphasize any cases where k is too small, in order to avoid the equilibrium being reduced to the boundary ones. The boundary equilibria have been discussed in Section 4.3.3.

In today's reality, with the help of the internet, information regarding asset management services has become more transparent. On the one hand, it is easier for households to acquire information concerning fund managers' investment principles and historical performance, which helps households to target competent managers more precisely. On the other hand, managers share more expertise and accessibility to discover the fundamental values of companies. The emergence of electronic trading and data mining has greatly reduced the friction for managers to trade and acquire information. Therefore, it can be anticipated that demand for factor investing is growing.

Lastly, the information frictions directly affect the level of opportunism in the active industry (equation (4.3.13)). Straightforwardly, demand for factor investing is higher with a higher level of opportunism in the active industry. This can be explained from the following two perspectives. First, due to the popularity of active management in the past half century, the funds engage in fierce competition with each other, and share high levels

of similarity in their strategies.[21] These lead to the decrease in the number of managers who can stand out and still deliver stellar returns compared to their peers. Second, active funds have long been believed to adopt skills enabling them to generate excess returns. But as more and more performance analysis and academic research emerge, households are now more aware of which funds are truly competent, and are also more convinced of the existence of free-riders in the industry. This trend could possibly lead to a smaller μ and more people opting for factor investing as a safer substitute.

Risk appetite

I examine how changes of risk appetite affect the asset market equilibrium. The model predicts that both factor investing and fundamental investing reach peak level at an intermediate risk aversion level, contradicting the argument in Glasserman and Mamaysky (2018).

As can be seen from Figure 4.4, the equilibrium capital allocations are of a concave and humped pattern, as the investors become more risk-averse. Funds in essence hold inside information concerning the volatility of either systematic or idiosyncratic loadings. When $\gamma \to 0$, the investors are inclined to be risk-neutral, whereby, expected returns become the key focus, while volatility and signals play only a very limited role. At the same time, asset prices become highly informative because the prices are more sensitive to supply shocks. Hence, the capital allocated to funds is diminishing and households instead prefer direct investment. When γ becomes too large, households are highly risk-averse, thus they gradually shift from risky assets to riskless bond, causing a decrease in fund delegation as well. Thus there exists an intermediate maximal of demand for delegation, where both conditions are satisfied: (i) households are willing to hold risky assets, and (ii) they care about risk, which is characterized by volatility. Note that the maximal demands for factor and fundamental investing are not reached at the same risk aversion level.

I also find that the fraction of informed households, ξ, is of a convex shape with respect to risk appetite. Information is more valuable when there are fewer skilled managers, typically, when risk-aversion is low. At a relatively high level of risk-aversion, households hold fewer risky assets. This is when the cost of investing in wrong risky assets becomes extremely high. Thus more households incur some searching cost to detect closet indexing. Moreover, though not shown in the figure, high risk aversion harms price efficiency. The decreasing price efficiency is consistent with the prediction in the Kyle (1985) model, where higher levels of risk aversion lead those who are informed to trade less, engendering a less informative price.

[21] See, e.g., Chevalier and Ellison (1999) and Hong et al. (2005).

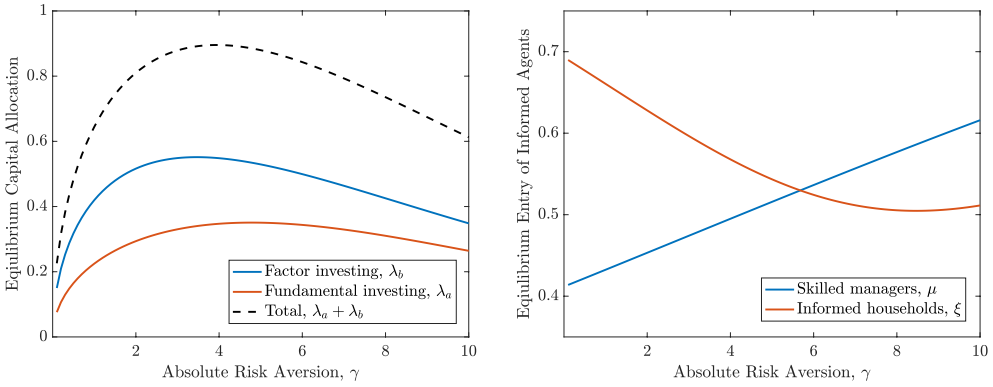


Figure 4.4: *Comparative statics w.r.t. investors' risk appetite.* As I vary the choice of γ in $(0, 10]$, this figure shows the equilibrium capital allocation, and amount of skilled managers and informed households. For the other parameters, take $k = 0.07$, $c = 0.1$, $\sigma^2_{q_I} = \sigma^2_{q_2} = \sigma^2_e = \sigma^2_s = 2\,\sigma^2_\varepsilon = 0.2$.

As households gradually become more aware of the assets, risks, and investment tools, they develop a more comprehensive view of wealth management. Instead of putting money into banks, they become more willing to hold risky assets. At the same time, they care about both risk and return. This development, in my model, can be translated into an intermediate level of absolute risk aversion γ.

Supply uncertainty

An unstable environment of individual assets, including return volatility, supply uncertainty and information accessibility, will affect people's delegation choices. If the asset-specific risk becomes higher (σ^2_e increases), or the information about idiosyncratic loadings becomes less precise (σ^2_ε increases), or the supply becomes highly uncertain ($\sigma^2_{q_2}$ increases), factor investing becomes more attractive, because systematic risk is relatively easy to control. For risk-averse households, factor investing delivers a safer and more stable payoff compared to fundamental investing. At the same time, fundamental investors can still attract greater demand because they partly profit from the idiosyncratic risks.

I take the supply uncertainty as an example. More noise trading leads to higher supply uncertainty of assets. As I vary the supply uncertainty of individual asset $\sigma^2_{q_2}$, the equilibrium capital allocation is shown in Figure 4.5. The figure does not cover the cases where the supply is too volatile. This is when demand for factor investing becomes non-monotonic, as it reaches the ceiling and has fully exploited the informational advantage over market factor.

A high supply uncertainty drives up the capital in the overall delegation industry, including both factor and fundamental investing, and curtails the delegation fees. Skilled

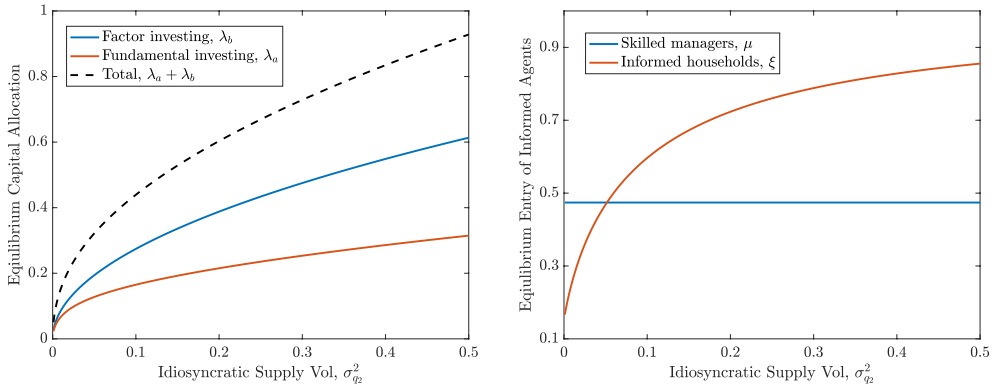


Figure 4.5: *Comparative statics w.r.t. supply uncertainty.* As I vary the choice of $\sigma^2_{q_2}$ in $(0, 0.5]$, this figure shows the equilibrium capital allocation, and amount of skilled managers and informed households. For the other parameters, take $k = 0.07,\ c = 0.1,\ \gamma = 3,\ \sigma^2_{q_I} = \sigma^2_e = \sigma^2_s = 2\,\sigma^2_\varepsilon = 0.1$.

managers do benefit from signaling their superiority because of an upward sloping curve of informed households. However, the decreasing fees deteriorate their gains directly. In aggregation, the amount of skilled managers remains unchanged. Interestingly, the impact of idiosyncratic supply uncertainty on capital allocation cancels out its effect on the price informativeness of idiosyncratic loading itself. Hence, when there is a supply shock, the size of the shock, $\sigma^2_{q_2}$, affects mainly the price efficiency of systematic loading. This is because part of the effect of idiosyncratic supply uncertainty is internalized by the market itself through altering the allocation preference of households.

4.5.1 Effects on return volatility and price efficiency

I now investigate the change of price terms when the capital in factor investing is larger. First I give some theoretical hints as to how to analyze the return volatility and price efficiency.

An increase in factor investing can be compensated by a decrease in either (i) fundamental investing or (ii) direct investing. Thus I define two ratios. The first ratio is the amount of factor investing to fundamental investing, i.e.,

$$r_{f-f} = \frac{\lambda_b}{\lambda_a}. \tag{4.5.24}$$

If the rise of factor investing does not affect direct investing much, yet curtails fundamental investing, then r_{f-f} would rise. The second ratio is the capital between non-fund users and fund users, i.e.,

$$r_{n-f} = \frac{\lambda_h}{\lambda_b + \lambda_a}. \tag{4.5.25}$$

If an increase in factor investing curtails direct investing, r_{n-f} would drop. Therefore, to examine the influence of increased factor investing, I examine how the price terms change with respect to the two ratios defined by (4.5.24) and (4.5.25).

Provided with the equilibrium price specified in Proposition 4.2.1, I look into its changes in value, stability, and co-movement brought by rising factor investing, as shown in Lemma 4.5.1.

Lemma 4.5.1 (**Factor investing and price**)

(i) If increased factor investing is compensated by less fundamental investing (r_{f-f} rises):
  (a) The price of asset-idiosyncratic component becomes more volatile;
  (b) The covariance between the prices of asset-idiosyncratic components increases;
  (c) The return of asset-idiosyncratic component becomes more volatile.

(ii) If increased factor investing is compensated by less direct investing (r_{n-f} drops):
  (a) The expected prices increase;
  (b) The index price becomes more volatile, if the inside information is precise enough;
  (c) The covariance between individual prices increases, if trading of the index portfolio is liquid enough.

The proof of Lemma 4.5.1 shows that the increasing instability, given by result (ia), is mainly attributed to the idiosyncratic component of an asset. The price could also be more volatile for the index in certain scenarios, as stated by (iib). Second, factor funds all feature the same bunch of portfolio, in which sense the investment market becomes highly concentrated through the popularity of factor investing. The model shows that a more concentrated industry is more vulnerable to idiosyncratic supply shock. Third, the rise in factor investing increases the price co-movement of assets in the index, given in (ib). In fact, one can tell from the proof that the influence of r_{f-f} on variance and covariance of assets is not linear but quadratic. The analysis lends support to the recent regulation reports, Sushko and Turner (2018) and Anadu et al. (2018), highlighting that fundamental-to-factor shift may affect financial stability. The co-movement result also lines up with the model extension outcome in Basak and Pavlova (2013).

Finally, the idiosyncratic risk premium becomes more volatile as λ_b increases. This is where active funds benefit from, in that the excess utility generated by fundamental over factor investing managers is positively correlated with the risk premium uncertainty.[22] Thus a byproduct of this conclusion is the increase in the outperformance of skilled active over factor funds. If the rising factor investing is accompanied by a decrease in fundamental investing, the remaining active managers benefit from the less competitive environment

[22]The argument comes directly from the results in Lemma D.3.1.

thus are more likely to generate superior performance. The intuition of this result is provided by a FCA occasional paper (James et al., 2019).

Increased factor investing has impact on the amounts of informed trading in different portfolios, which, for sure, will affect the information embedded into asset prices.

Lemma 4.5.2 (**Factor investing and price efficiency**)

(i) If increased factor investing is compensated by less fundamental investing (r_{f-f} rises), the price efficiency of individual assets decreases (represented by an increasing η_i, $i = 1, ..., l$);

(ii) If increased factor investing is compensated by less direct investing (r_{n-f} drops), the price efficiency of market index increases (represented by a decreasing η_I).

The influence of r_{f-f} on the information efficiency is neither linear nor quadratic, yet of a more complex relation as stated in the proof. Internalizing the amount of factor investing, I summarize a few possible changes that can harm price efficiency in Figure 4.6. I recover from definition (4.2.6) and (4.2.7) the original price efficiency measures, i.e., $\frac{Var(S|s)}{Var(S|p_I)}$ for market factor and $\frac{Var(v_2|\mathcal{I}_a)}{Var(v_2|p_2)}$ for individual asset.

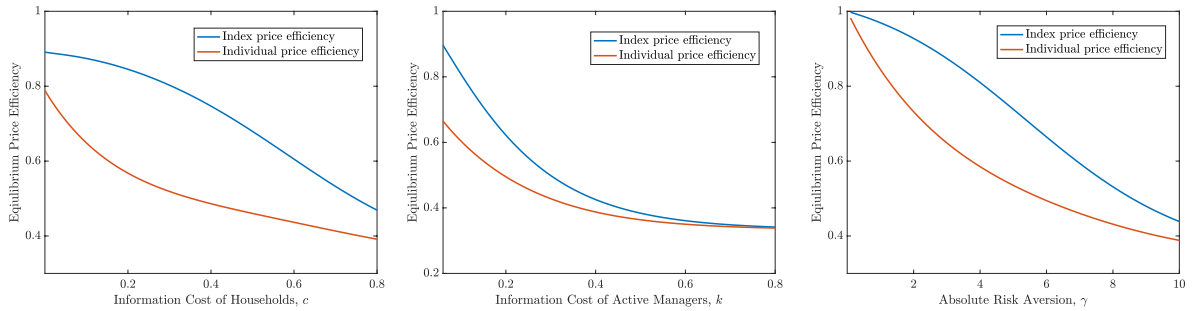


Figure 4.6: *Exogenous factors harming price efficiency.* As I vary the choice of $c \in [0, 0.8]$, or $k \in (0, 0.8]$, or $\gamma \in (0, 10]$, this figure shows the corresponding effect on price efficiency. Assume the individual stock has an average exposure $\beta_2 = 1$. For the other parameters, take $\gamma = 3$, $k = 0.07$, $c = 0.1$ if they are fixed variables, $\sigma^2_{q_I} = \sigma^2_{q_2} = \sigma^2_e = \sigma^2_s = 2\,\sigma^2_\varepsilon = 0.2$.

The model predicts that high information frictions or high risk aversion deteriorate the security-specific information and systematic information contained in prices. The outcome partly contradicts the argument in Cong and Xu (2016), in which the authors introduce factor investing exogenously. They argue that the rising factor investing harms idiosyncratic efficiency. In this model, this is true when the searching cost of households to find out manager type is tiny or when households are inclined to be risk-neutral. Then any increase in households' searching cost or in their risk aversion will give rise to factor investing, curtailing the price efficiency of assets. However, when c or γ further increases,

or the fundamental information is harder for managers to obtain, there is a decreasing demand for factor investing, accompanied by decreasing price efficiency. This contradicts the outcomes in Cong and Xu (2016). Moreover, the amount of fund delegation first increases with risk aversion but then decreases, whereas price efficiency monotonically decreases. This is because price informativeness is related to the risk-appetite-adjusted delegated capital. With investors more inclined to be risk-neutral, the two effects cancel each other out.

4.6 Conclusion

In this paper I have developed a general equilibrium, jointly determining the households' capital allocation, their investment in finding out manager quality, the roles of fund managers, and fund fees. The outcomes contribute to understanding the segmentation of wealth under different investment vehicles, including self-investing and different types of funds.

The model also helps answer the regulatory concerns about the impact of rising factor investing on financial stability. Studies like Malikov (2019) imply that public concerns are merited as, according to his framework, the increase in indexing could be unbounded. However, in this paper, the equilibrium implies an implicit ceiling of capital in factor investing. Upon reaching this level, indexation has fully exploited its informational advantage in relation to market factor. Therefore, regulators do not need to step in to prevent factor investing from seizing 100 percent of the market.

Furthermore, the model bridges the standard REE literature with household finance. The model shows that household inequality in their wealth endowments and in their cognitive ability to detect manager quality serves as one possible explanation of the rise in factor investing. The wealth effect, in this framework, is more prominent than the cognitive ability effect. This novel result encourages future empirical testing of the links between wealth inequality and capital allocation shifts. Other avenues for future research involve looking at innovative ways of modeling cognitive ability, for instance by incorporating its impact on certainty-equivalent utility. In addition, the model claims that the falling but still not insignificant information costs, an intermediate level of household risk aversion, or high idiosyncratic risk and supply uncertainty can also give rise to factor investing.

Lastly, the model has certain implications on searching for the optimal industrial organization solution for the delegation industry. For the active industry to guarantee flows, in any equilibrium there cannot be an excessive number of closet indexers. Meanwhile managers cannot set management fees too high, as factor investing brings additional competitive

pressure to the delegation industry. The model also helps understand empirical asset pricing findings related to financial intermediaries, including return volatility, co-movement, and price efficiency.

Chapter 5

Conclusion

In this thesis, we study the effects of incentives, contracts, information, and liquidity mismatch on the asset management industry.

The welfare of asset management professionals frequently depends both on tournament rewards for relative performance and compensation-based absolute performance rewards. In Chapter 2, we study the mixed incentives and risk-taking strategies arising from combining the two types of rewards. Surprisingly we find that introducing bonus rewards into relative performance rewards can reduce downside risk-taking, sometimes even upside risk-taking. This is because bonus rewards distract superior competitors from rank competition, thus increase weak contestant's gains from rank competition, thereby leveling the playing field and consequently reducing downside risk taking by weak competitors. The effect is founded on very generic properties of rank contests: it persists in dynamic, asymmetric, multi-contestant, or risk-averse settings.

Therefore, mixed incentives can engender less overall risk taking than either pure rank or pure bonus incentives. The effects partly echo the long-term search for socially beneficial compensation packages. The findings also shed light on the real-world evidence and policy implication: (i) Curtailing high-powered absolute performance may sometimes increase risk-taking. (ii) Agents receiving mixed compensation, as can be widely observed in asset management industry, basically eschew extremely risky strategies. (iii) High-powered bonus is common and generous, especially when the agents' performance is publicly available.

In Chapter 3, we address a policy concern of fragility in open-end funds, which poses a significant threat to economic stability and social welfare. Existing literature helps us understand the origin of runs on open-end funds—the traditional pricing rules that lead to dilution effect and thus first-mover advantage. In this study, we investigate how to mitigate the run risk during stress periods. Through analyzing the alternative pricing rules which adjust funds' NAV to pass on funds' trading costs to transacting shareholders,

we find that the new rules indeed help funds retain capital under stress market conditions. The new rules thus serve as substitutes to other means of liquidity risk management. For instance, the alternative funds have less cash holdings. This stabilizing effect is internalized particularly by institutional investors and investors with longer investment horizons, those who alter their behavior the most after the funds switch from traditional to alternative pricing rules.

However, we also identify that funds using the new rules have difficulty attracting investor capital outside stressed periods. This is mostly associated with tracking error: the swinging prices bring unnecessary volatility irrelevant to the changes in fundamental value. The drawback of the alternative pricing rules partly explains the phenomenon that not all the funds choose to adopt the new rules.

In Chapter 4, I build a theoretical framework to link wealth inequality to the pertinent features in asset management industry. Typically, I show that wealth inequality has two effects: (i) increases demand for indexing, and (ii) forces both active and index funds to lower fees. Because of the existence of "closet indexers" in active industry, i.e., those managers who market themselves as stock-pickers yet hold massively index-like portfolios, skilled active funds would attract the wealthiest households. Faced with a fixed cost to detect closet indexing, wealthier individuals have relatively cheaper information cost, therefore are more willing to figure out the manager type. Therefore, active managers have to compete more aggressively through information acquisition. More information production brings individual prices closer to the fundamentals, thereby lowers active funds' excess returns and the fees they charge. The model provides justification for the assumption of decreasing returns to scale both at fund level (e.g., Berk and Green, 2004) and at industry level (e.g., Pástor and Stambaugh, 2012).

Furthermore, the equilibrium outcome of household capital allocation suggests that there exists an upper bound of the capital in index funds. Therefore, even though we observe a remarkable growth of assets under index funds, the regulators and practitioners need not worry about a 100% indexing in the market. More importantly, the exact level of this upper bound can be measured directly using the amount of closet indexers, which, with feasible measures in the existing studies, can also be identified.

Appendix A

Notations

Table A.1: *Frequently used notations and variable definitions.*

Notation	**Description**
Chapter 2	
S, W	Strong and weak managers
X_S, X_W	Random performance at date 1 of manager S, W
b	Bonus reward if exceeding hurdle
r	Threshold which triggers bonus payment
μ_S, μ_W	The capacity of manager S, W $(0 < \mu_W \leq \mu_S)$
u	Upper bound of the support of Eq1 equilibrium distribution
u_L, u_H	Upper bound of the support in pure rank, and rand and bonus competition region of Eq2 equilibrium distribution
$F_S(\cdot), F_W(\cdot)$	Challenge performance distributions picked by manager S, W
$\Pi_S(\cdot), \Pi_W(\cdot)$	Payoff functions (expected rewards) of manager S, W
α_S, α_W	Intercept of manager S, W's support line
β_S, β_W	Slope of manager S, W's support line
α'_S, α'_W	Intercept of manager S, W's support line under linear compensation
β'_S, β'_W	Slope of manager S, W's support line under linear compensation
$F^*_S(\cdot), F^*_W(\cdot)$	The equilibrium performance distributions chosen by manager S, W
p^r_S	Point mass at the bonus threshold r placed by the strong manager
p^0_W	Point mass at 0 placed by the strong manager
p^h_S	Probability of the strong manager to choose a uniform distribution over $[r, u_H]$
p^h_W	Probability of the weak manager to choose a uniform distribution over $[r, u_H]$
F, G	Two managers/contestants without specifying strong or weak position

(*Continued*)

ι_F, ι_G	The weight of manager F, G putting on the bonus reward incentive
μ_F^L, μ_G^L	Manager F, G's conditional capacity allocated to the pure rank competition region
μ_F^H, μ_G^H	Manager F, G's conditional capacity allocated to the rank and bonus competition region
p_F, p_G	Manager F, G's probability of targeting the pure rank competition region
$N(\cdot, \cdot)$	Rank reward winning function, piecewise, differentiable and quasilinear
t	The time period in the dynamic game, $t \in \{0, 1, 2, ...\}$
δ	The periodic gain/loss of capacity when a manager win/loss in the dynamic game
ϕ	Discount factor applied to continuation values, $\phi \in (0, 1)$
$V_F(\cdot, \cdot), V_G(\cdot, \cdot)$	The value of contestant F, G in the dynamic game (for simplification, we drop subscript when analyzing a single manager)
$V_F^*(\cdot, \cdot), V_G^*(\cdot, \cdot)$	The equilibrium value (fixed point) of contestant F, G in the dynamic game (for simplification, we drop subscript when analyzing a single manager, and use subscript 'L' and 'T' to represent leading and trailing contestant)
Π_F, Π_G	Reward in the current period of manager F, G in the dynamic game
$u(\cdot)$	CARA utility function of risk-averse contestants
θ	Risk aversion parameter in the CARA utility function
$\mathcal{L}$	Lagrangian function in static game
$\mathcal{L}'$	Lagrangian function at a given state in dynamic game
λ_F, λ_G	Lagrangian multipliers of contestant F, G
Θ_F, Θ_G	Policy vector, $\Theta_i \coloneqq \{p_i, \mu_i^L, \mu_i^H\} \in [0, 1] \times [0, \infty) \times [r, \infty)$, $i = F$ or G
ω	State in the dynamic game, $\omega \in \Omega$
$\Gamma(\cdot)$	Transition function, mapping the current state into the state in the following period
$T[\cdot](\cdot)$	Bellman equation
v_0	Starting value of iteration in contraction mapping
Chapter 3	
Stress	An indicator variable that equals one if monthly VIX index is above the 75^{th} percentile of the sample
Alternative	An indicator variable that equals one if the fund is using one of the alternative pricing rules

(*Continued*)

Flow	Monthly capital flows into a fund divided by fund's total net assets in t (%)
Flow EndInv	Change in each investor's holding (in number of shares) from previous month (%)
Return	Fund's monthly raw return (%)
Alpha	Estimated using rolling-window time-series regression for each fund using the past 12 months data. Alpha is the intercept from a regression of excess fund returns on excess global bond market and global stock market returns. Indices obtained from Barclays (%)
NegAlpha	Equals Alpha if the fund's Alpha is negative (or below the 25^{th} percentile); set to zero otherwise (%)
Size	Natural logarithm of fund's total net assets (£)
Age	Natural logarithm of fund age in years (using the age of the oldest class share)
Expense	Fund's annual total expense ratio (%)
N of Inv	Natural logarithm of total number of investors in a given fund
Illiquidity	Value-weighted average of Asset Illiquidity of fund's assets
Asset Illiquidity	Bid-ask spread; end of day bid and ask prices are obtained from Thomson Reuters Datastream and used in the following order depending on availability: Thomson Reuters composite price, Thomson Reuters Pricing Service evaluated price, iBOXX, and ICMA
Inst	Fraction of fund's total net assets held by institutional investors (%)
Ownership Concentration	Herfindahl-Hirschman Index calculated using each end-investors' ownership in each month
Adjustment Factor	Equals the absolute value of swing factor for swing funds; equals half-spread, $0.5 * (ask - bid)/mid$, for dual funds (%)
Net Inflow	Net monthly inflows. Equal to Flow if Flow>0; equal to 0 if Flow≤ 0
Net Outflow	Net monthly outflows. Equal to Flow if Flow<0; equal to 0 if Flow≥ 0
Dual	An indicator variable that equals one if the fund is a dual fund
Full	An indicator variable that equals one if the fund is a full swing fund
Partial	An indicator variable that equals one if the fund is a partial swing fund
Cash	Fund's total cash holding—defined as cash plus cash equivalents including cash deposits, money market funds, Treasury Bills, commercial paper, short term bonds, repos and currency holdings—divided by the value of total assets (%)

(*Continued*)

Tracking Error	-1 times R-squared from the alpha regression described above
Asset Conc	Herfindahl–Hirschman Index of fund's asset holdings in each month
Front Load	Value-weighted average of (minimum) front-end load charges across share classes of a given fund (%)
Back Load	Value-weighted average of (minimum) back-end load charges across share classes of a given fund (%)
New Investor	Number of new investors divided by the fund's total number of investors in each month (%)
Investor Horizon	Number of months the investor holds his shares after an initial purchase. We use purchases before December 2014
Chapter 4	
l	Number of risky assets
h	Households investing directly
a	Households employing active funds
b	Households employing factor funds
$\boldsymbol{v} = (v_1, ..., v_l)^\intercal$	Payoffs of risky assets
$\boldsymbol{x} = (b_1, ..., b_l)^\intercal$	Index portfolio
$\boldsymbol{p} = (p_1, ..., p_l)^\intercal$	Equilibrium price of assets
$\boldsymbol{q}$	Aggregate market supply, $\mathcal{N}(\overline{\boldsymbol{q}}, \Sigma_q)$ distributed
S	Payoff of factor loading, $\mathcal{N}(\overline{s}, \sigma_s^2)$ distributed
$\mathcal{E}_i$	Payoff of idiosyncratic loading, $\mathcal{N}(0, \sigma_e^2)$ distributed, $i \in \{1, 2, ..., l\}$
β_i	Asset i's exposure to the aggregate market, $i \in \{1, 2, ..., l\}$
$\boldsymbol{x}_j$	Portfolio holding strategy of investors, $j \in \{a, b, h\}$
u_j	Certainty-equivalent wealth of investors, $j \in \{a, b, h\}$
$\mathcal{I}_j$	Information set of investors, $j \in \{a, b, h\}$
λ_j	Households' capital allocation strategy, $j \in \{a, b, h\}$
α_j	Fund delegation fees, $j \in \{a, b\}$
$\boldsymbol{\Sigma}_a$	Variance-covariance matrix conditional on active fund managers' information set
s	Imperfect signal of systematic loading, with noise $\varepsilon_0 \overset{d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$
$\boldsymbol{e} = (e_1, ..., e_l)^\intercal$	Imperfect signals of idiosyncratic loadings, with noise $\varepsilon_i \overset{d}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, $i \in \{1, 2, ..., l\}$
p_I	Price of index portfolio
q_I	Market supply of index portfolio, $\mathcal{N}(\overline{q}_I, \sigma_{q_I}^2)$ distributed

(*Continued*)

q_i	Market supply of the rest orthogonal assets, $\mathcal{N}(\overline{q}_i, \sigma^2_{q_i})$ distributed, $i \in \{2, ..., l\}$
η_I	Price inefficiency of index portfolio
η_i	Price inefficiency of individual asset i, $i \in \{1, 2, ..., l\}$
n	A continuum of households (household n with superscript n)
γ	Absolute risk aversion
γ^R	Relative risk aversion
W	Initial wealth or endowment
k	Managers' cost of acquiring fundamental information
c	Households' cost of searching and vetting a manager
μ	Fraction of skilled active managers
ξ	Fraction of informed households
$B(a, b)$	Beta function with parameter (a, b)
$\Gamma(a, b)$	Gamma function with parameter (a, b)
$AMH(\theta)$	Ali-Mikhail-Haq copula with parameter θ
r_{f-f}	The ratio of factor investing share to fundamental investing share
r_{n-f}	The ratio of non-fund users' capital to fund users' capital

Abbreviation	**Description**
cf.	*confer/conferatur*
e.g.	*exempli gratia*
etc	*et cetera*
i.e.	*id est*
AUM	Asset under management
CAPM	Capital Asset Pricing Model
CARA	Constant absolute risk aversion
CRRA	Constant relative risk aversion
CDF	Cumulative distribution function
ECB	European Central Bank
ERISA	Employee Retirement Income Security Act
ETF	Exchange-Traded Funds
FCA	Financial Conduct Authority
FE	Fixed effect
GDP	Gross domestic product
GP	General partner
IPO	Initial public offering
IRR	Internal rate of return

(*Continued*)

LHS	Left hand side
LP	Limited partner
LPA	Limited Partnership Agreement
MSA	Management Services Agreement
NAV	Net asset value
OLS	Ordinary least squares
PDF	Probability density function
REE	Rational expectations equilibrium
RHS	Right hand side
SEC	U.S. Securities and Exchange Commission
SML	Security market line
TNA	Total net assets
VC	Venture capital
VIX	Chicago Board Options Exchange Volatility Index

Appendix B

Appendix: Atalanta Effect: How High-Powered Compensation Reduces Risk-Taking

Abstract: This chapter contains proofs of the results (Section B.1), the analysis of contest equilibria with more than two contestants (Section B.2), the details on the construction of the numerical solution to the dynamic contest problem (Section B.3), the formal derivations of the generic properties of contest equilibria (Section B.4), the configurations of easy bonus equilibria (Section B.5), the designer welfare analysis (Section B.6), and the algorithm of finding an optimal compensation package from the contest designer's perspective (Section B.7).

B.1 Proofs of Results

Proof of Lemma 2.3.1 Lemma 2.3.1 is a consequence of the lemma proved below. However, in order to permit the reader to follow the derivations of the results in the sequence in which they are presented in the manuscript, we provide a self-contained derivation below. The derivation is in the spirit of the derivation used in Hillman and Riley (1989), but their derivation is less detailed and their model is a bit different than ours. This derivation will exploit the fact that the pure rank competition game is a two-player constant sum game. Thus, each contestant's payoff is maximized by minimizing the payoff of the rival contestant. Consequently, the game can be solved using minimax approach.

A necessary (but not sufficient) condition for an equilibrium in which neither contestant chases the bonus is that the performance distributions of the contestants are equilibrium distributions for a pure rank competition contest, i.e., a contest where $b = 0$. To see this, note that the payoff for the contestants if neither chases the bonus, equals their payoff in a pure rank competition. If an alternative performance distribution produced a higher payoff, even ignoring whatever rank rewards it might garner, then a candidate

equilibrium performance distribution, that candidate equilibrium distribution could not be a best reply.

We will show that the pure rank competition game has a unique equilibrium which yields the equilibrium performance distributions specified in Lemma 2.3.1. Thus, these performance distributions must characterize equilibrium performance when neither contestant chases the bonus. Let i be a given contestant and let j be the rival contestant. The definitions of the payoff function given by equation (2.3.3), show that, for a given performance distribution for i, F_i, j's payoff equals

$$\max_{dF_j}\{\int_0^\infty F_i(x)\,\mathrm{d}F_j(x) : \int_0^\infty x\,\mathrm{d}F_j(x) \leq \mu_j \;\&\; F(0-) = 0.\} = \hat{F}_i(\mu_j),$$

where $\hat{F}_i$ is the concave upper envelope of F_i, i.e., $\hat{F}_i$ is the pointwise infimum of all affine functions that majorize F_i. This assertion simply follows from the definition of the concave upper envelope (cf. Aumann et al., 1995; Gentzkow and Kamenica, 2014). Because the total rank rewards received by i and j always sum to 1, the payoff to i equals $1 - \hat{F}_i(\mu_j)$. Thus, i's payoff is maximized by minimizing $\hat{F}_i(\mu_j)$ over feasible performance distributions.

Because, by definition, the concave upper envelope of F_i is a concave distribution that is stochastically dominated by F_i, $\hat{F}_i$ is feasible whenever F_i is feasible. If F_i is not concave, then F_i would strictly stochastically dominate $\hat{F}_i$. In which case, if F_i is feasible, $\hat{F}_i$ would produce the same payoff as F_i and strictly satisfy the capacity constraint. Thus, F_i would not be optimal because a sufficiently small adjustment in $\hat{F}_i$ would reduce the probability of the rival contest j winning and still satisfy the capacity constraint.

Thus, all optimal performance distributions for i are concave. Because the integral of the survival function, $S_i = 1 - F_i$, of a non-negative random variable equals its expectation, it is convenient to re-express these observations in terms of survival functions rather than distribution functions. Concavity of the distribution function implies convexity of the survival function. Expressed in terms of survival functions, i's optimal survival functions are the survival functions which solve,

$$\text{PS:}\quad \max_{S_i}\{S_i(\mu_j) : S_i \text{ is a convex survival function, } S_i(0-) = 1 \;\&\; \int_0^\infty S_i(x)\,\mathrm{d}x \leq \mu_i\}.$$

Next we show, by a proof by contradiction, that all solutions to PS are affine over their support: suppose that there exists a solution, S_i^o, that is not affine over its support. Let ℓ be an affine function that is a lower support line for S_i^o at μ_j. Convexity assures the existence of ℓ. Then, by definition, $\ell(\mu_j) = S_i^o(\mu_j)$, and $\ell \leq S_i^o$. Let $S_i'(x) = \max[\ell(x), 0]$. Then S_i' is a convex survival function and $S_i' \leq S_i^o$ and $S_i'(\mu_j) = S_i^o(\mu_j)$. Because S_i^o is not

affine, $S_i' < S_i^o$ on a set of positive Lebesgue measure; thus,

$$\int_0^\infty x\,\mathrm{d}F_i'(x) = \int_0^\infty S_i'(x)\,\mathrm{d}x < \int_0^\infty S_i^o(x)\,\mathrm{d}x = \int_0^\infty x\,\mathrm{d}F_i(x) \leq \mu_i.$$

Thus, the capacity constraint is not binding at S_i'. Hence, it is possible to find a feasible survival function, S_i'' such that

$$S_i''(\mu_j) > S_i^o(\mu_j) \text{ and } \int_0^\infty S_i''(x)\,\mathrm{d}x \leq \mu_i, \tag{B.1.1}$$

through a small increase in the slope or intercept of ℓ. Equation (B.1.1) contradicts the optimality of the candidate non-affine solution, S^o. This argument is encapsulated by Figure B.1.

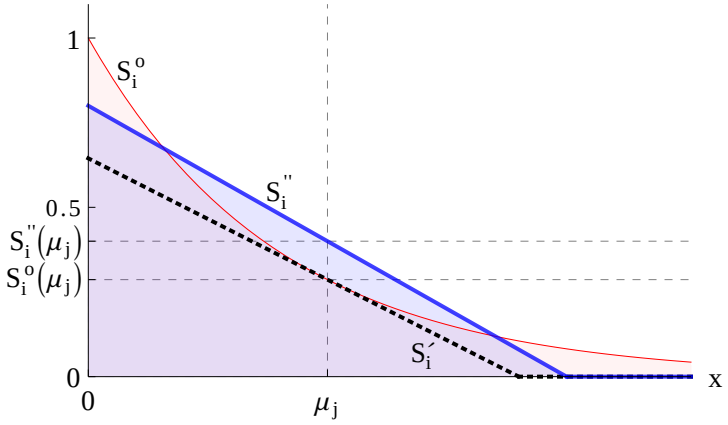


Figure B.1: *Optimal survival functions in pure rank competitions.*

Thus, optimal solutions to PS are affine over their supports and hence have the form, $S(x) = \max[a - b\,x, 0]$, $a \in [0,1]$ and $b > 0]$. Noting that

$$\int_0^\infty \max[a - b\,x, 0]\,\mathrm{d}x = \frac{a^2}{2\,b},$$

we see that optimal survival functions solve the following simple programming problem:

$$\text{PSAff}: \quad \max_{a\in[0,1],b>0} \{a - b\mu_j : \frac{a^2}{2\,b} \leq \mu_i\}.$$

The unique solution to PSAff is given by

$$S_i^*(x) = \min\left[\frac{\mu_i}{\mu_j}, 1\right] - \frac{1}{2\,\mu_j} \min\left[\frac{\mu_i}{\mu_j}, 1\right] x.$$

Thus, in any equilibrium in which neither S nor W chase the bonus, S^* characterizes optimal performance distributions. Considering the case where $i = S$ and $j = W$ as well

as the case where $i = W$ and $j = S$, and translating the survival functions in to distribution functions, yields the characterizations of contestant strategies in Lemma 2.3.1.

Proof of Lemma 2.3.2 Suppose, to obtain a contradiction, that there exists an equilibrium in which only W chases the bonus. Thus, r is in the support of W's performance distribution. By Remark 2.3.3, the reward from choosing $x = r$, $1 + b$, lies on W's support line, i.e.,

$$\alpha_W + \beta_W\, r = 1 + b. \tag{B.1.2}$$

Similarly, because not chasing the bonus is a best response for S, the payoff to S from not chasing the bonus lies weakly below S's support line, i.e.,

$$\alpha_S + \beta_S\, r \geq 1 + b. \tag{B.1.3}$$

Equations (B.1.2) and (B.1.3) imply that

$$\alpha_S + \beta_S\, r \geq \alpha_W + \beta_W\, r. \tag{B.1.4}$$

Now let $\bar{x} = \max(\text{Supp}_S)$. The capacity constraint implies that $\bar{x} > 0$. The hypothesis that S is not chasing the bonus implies that $\bar{x} < r$. For $x < r$, the reward to S for performance x, $\Pi_S(x) = F_W(x)$, and, similarly, $\Pi_W(x) = F_S(x)$. Moreover, over $[0, r)$, the supports of F_S and F_W coincide (Remarks 2.3.2.(a) and 2.3.2.(d)) and, hence, $\bar{x} \in \text{Supp}_S \cap \text{Supp}_W$. This implies that $\bar{x}$ lies on both S's and W's support lines. Because W is chasing the bonus and S is not, $F_W(\bar{x}) < 1 = F_S(\bar{x})$. These observations imply that

$$\alpha_S + \beta_S\, \bar{x} = F_W(\bar{x}) < 1 = F_S(\bar{x}) = \alpha_W + \beta_W\, \bar{x},$$

which, in turn, implies that

$$\alpha_S + \beta_S\, \bar{x} < \alpha_W + \beta_W\, \bar{x}, \quad \bar{x} < r. \tag{B.1.5}$$

Now note that, because S's capacity exceeds W's, it cannot be the case that the performance distribution of W first-order stochastically dominates the performance distribution of S. Because S is not chasing the bonus, this implies that there must exist at least one point, say x_o such that $x_o \in [0, \bar{x})$ and $F_W(x_o) > F_S(x_o)$. Using the same argument as used in the previous case, we see that

$$\alpha_S + \beta_S\, x_o = F_W(x_o) > F_S(x_o) = \alpha_W + \beta_W\, x_o,$$

which implies that,

$$\alpha_S + \beta_S\, x_o > \alpha_W + \beta_W\, x_o, \quad x_o \in [0, \bar{x}). \tag{B.1.6}$$

Inspection shows that equations (B.1.4), (B.1.5), and (B.1.6) cannot be simultaneously satisfied for any choice of multipliers, α_S, β_S, α_W, β_W, and thereby establishes the contradiction.

Proof of Lemma 2.3.3 First note that, in an Eq1 equilibrium, because, by hypothesis, S is chasing the bonus and W is not, if S submits performance r, S wins both the rank reward of 1 and bonus reward of b. Performance $x > r$ will not produce a larger reward and requires more capacity. Thus, r is in the support of S's performance distribution, F_S, and r is not in the support of W's performance distribution, F_W, and $x > r$ is not in the support of either contestants' distributions.

We claim that the support lines for the two contestants will have the following properties: $\alpha_S > 0$, $\alpha_W = 0$, $\beta_S > 0$, and $\beta_W > 0$. The fact that β_S and β_W are positive follows simply from the fact that the marginal value of capacity, which can always be applied to either increasing the probability of winning the rank reward or capturing the bonus, is positive. Now consider α_S, and α_W.

First, note that it cannot be the case that both α_S and α_W are positive. Over the rank competition region, the reward to each contestant from performance x equals the distribution function selected by the other contestant evaluated at x. Let $\ell_i(x) = \alpha_i + \beta_i\, x$, $i = S, W$ represent the support lines of the contestants. Because 0 is in the support of both contestants' performance distributions (Remark 2.3.2.(d)), at point x in the support of the performance distribution, the reward to a contestant from performance x lies on the contestant's support line, i.e., $\ell_i(0) = \Pi_i(0) = \alpha_i$, $i = S, W$ (Remark 2.3.3). Thus, if both α_S and α_W were positive, $\Pi_S(0) > 0$ and $\Pi_W(0) > 0$. Since $\Pi_S(0) = F_W(0)$ and $\Pi_W(0) = F_S(0)$, this would imply that there is a positive probability of tied performance, which is impossible (Remark 2.3.1). So, if it is not the case that $\alpha_S > 0$ and $\alpha_W = 0$, then $\alpha_S = 0$ and $\alpha_W \geq 0$.

So to obtain a contradiction, suppose that $\alpha_S = 0$ and $\alpha_W \geq 0$. By hypothesis, only S chases the bonus. Thus, the bonus threshold, r, must exceed $\bar{x} = \max(\text{Supp}_W)$. Performance by S equal to $\bar{x}$ wins the rank competition with probability 1. But, because, by hypothesis, S is chasing the bonus, performance by W equal to $\bar{x}$ does not win the rank competition with probability 1. Because $\bar{x}$ is in the support of W's performance distribution, $\bar{x}$ is in the support of S's performance distribution (Remark 2.3.2.(a)). Hence, $\bar{x}$ lies on the contestant S's support line, ℓ_S (Remark 2.3.3); thus, $\ell_S(\bar{x}) = \Pi_S(\bar{x}) = 1 > \ell_W(\bar{x}) = \Pi_W(\bar{x})$. Because $\ell_W(0) \geq \ell_S(0)$, and $\ell_W(\bar{x}) < \ell_S(\bar{x})$, it must be the case that the slope of ℓ_S exceeds the slope of ℓ_W. Thus, at all $x > \bar{x}$, $\ell_S(x) > \ell_W(x)$. This implies, because

$r > \bar{x}$, it must be the case, that $\ell_S(r) > \ell_W(r)$. If only S chases the bonus, the reward from submitting performance r equals $1 + b$. Because, by hypothesis, S is chasing the bonus, r must lie on S's support line (Remark 2.3.3). So, $\ell_S(r) = 1 + b$. Thus, $\ell_W(r) < 1 + b$. The reward to W from submitting any performance greater than r is $1 + b$. Thus,

$$1 + b = \Pi_W(x) > \ell_W(r), \quad x > r.$$

Because the reward function is right continuous and the support line is continuous, there exists $x > r$ such that $\Pi_W(x) > \ell_W(x)$, which is inconsistent with the support line ℓ_W, being an upper bound on the reward function, and thus is inconsistent with W not chasing the bonus.

Consequently, in any Eq1 equilibrium $\alpha_W = 0$ and $\alpha_S > 0$. Because $\alpha_S = F_W(0)$ and $\alpha_W = F_S(0)$ this implies that W, and only W, places some probability mass on 0. Let p^0_W be the weight that W places on 0. Because a performance of r is sufficient to capture both the rank and bonus rewards, capacity has a positive shadow price, and, by hypothesis, S chases the bonus, the only point in the support of S performance distribution that is larger than the maximum performance of W is r. Thus, S places probability mass on r. Let p^r_S represent this weight. Let $u = \max(\text{Supp}_W)$ represent the upper bound on the performance of W. All $x \neq r$ are either in the supports of both contestants' distribution or in the supports of neither. Except perhaps at 0 and r, both equilibrium performance distributions are continuous (Remark 2.3.2.(c)). Moreover, Supp_W is connected (Remark 2.3.2.(b)) and the support of S's distribution, is the same as W's support except at $x = r$ (Remark 2.3.2.(a)). Thus, the support of W's performance distribution equals $[0, u]$ and the support of S's performance distribution equals $[0, u] \cup \{r\}$. All points in the support of the contestants' distributions lie on their support lines, which are affine. Thus, conditioned on choosing a performance level $x \in (0, u]$, both contestants submit uniformly distributed performance.

Thus, we know that in the equilibrium, contestant S puts point mass on r with probability p^r_S and plays a uniform distribution Unif$[0, u]$ with probability $1 - p^r_S$; whereas contestant W randomizes between 0 with probability p^0_W and uniform Unif$[0, u]$ with probability $1 - p^0_W$. Our next task is to relate these parameters to the contestant's capacities and the structure of bonus compensation, defined by r and b.

For contestant S, her support line intersects $(r, 1 + b)$, i.e.,

$$\alpha_S + \beta_S\, r = 1 + b. \tag{B.1.7}$$

Meanwhile if S submits u, she wins with probability one, i.e.,

$$\alpha_S + \beta_S\, u = 1. \tag{B.1.8}$$

Combining equation (B.1.7) and (B.1.8) we have $\beta_S = b/(r-u)$, $\alpha_S = 1 - (b\,u/(r-u))$, showing that

$$p_W^0 = \alpha_S = 1 - \frac{b\,u}{r-u}. \tag{B.1.9}$$

For contestant W, we require that

$$(1 - p_S^r)\,\frac{x}{u} = \beta_W\,x, \qquad x \in [0, u], \tag{B.1.10}$$

Hence, $\beta_W = (1 - p_S^r)/u$.

Because both contestants use up their capacities in equilibrium, we have

$$(1 - p_S^r)\,\frac{u}{2} + r\,p_S^r = \mu_S, \tag{B.1.11}$$

$$(1 - p_W^0)\,\frac{u}{2} = \mu_W. \tag{B.1.12}$$

Combining three equations (B.1.9), (B.1.11) and (B.1.12), we are able to solve for the three unknowns $\{u, p_S^r, p_W^0\}$. Given nonnegativity, there is only one solution:

$$u = \frac{2\,r\,\mu_W}{\mu_W + \sqrt{2\,b\,r\,\mu_W + \mu_W^2}}, \qquad p_S^r = \frac{2\,\mu_S - u}{2\,r - u}, \qquad p_W^0 = 1 - \frac{2}{u}\,\mu_W.$$

Finally, from the results above it is not hard to calculate the contest reward functions for both contestants. The reward function for contestant S is given by

$$\Pi_S(x) = \begin{cases} p_W^0 + (1 - p_W^0)\,\dfrac{x}{u} & \text{for } x \le u, \\ 1 & \text{for } x \in (u, r), \\ 1 + b & \text{for } x \ge r. \end{cases} \tag{B.1.13}$$

Correspondingly, the reward function for contestant W is given by

$$\Pi_W(x) = \begin{cases} (1 - p_S^r)\,\dfrac{x}{u} & \text{for } x \le u, \\ 1 - p_S^r & \text{for } x \in (u, r), \\ 1 + b & \text{for } x \ge r. \end{cases} \tag{B.1.14}$$

Moreover, the parameters of support lines satisfy

$$\alpha_S = 1 - \frac{b\,u}{r-u}, \qquad \beta_S = \frac{b}{r-u}, \qquad \beta_W = \frac{1 - p_S^r}{u}.$$

Proof of Proposition 2.3.1 As shown in Lemma 2.3.1, in Eq0,

$$p_W^0(\text{Eq0}) = 1 - \frac{\mu_W}{\mu_S}.$$

As shown in Lemma 2.3.3, in Eq1,

$$p_W^0(\text{Eq1}) = 1 - \frac{2}{u}\,\mu_W.$$

Lemma 2.3.3 shows, $u < 2\,\mu_S$, and thus $p_W^0(\text{Eq1}) < p_W^0(\text{Eq0})$.

Proof of Lemma 2.3.4 First note that, a proof analogous to the proof given for Eq1, shows that, in any Eq2 equilibrium, $\alpha_S > 0$ and $\alpha_W = 0$. Since, in an Eq2 equilibrium, both W and S chase the bonus, the supports of their performance distributions are identical. Let $\text{Supp} = \text{Supp}_W = \text{Supp}_S$. Let $u_H = \max(\text{Supp})$. Because u_H lies on the support of both contestants support lines, $\ell_S(u_H) = \Pi_H(u_H) = 1{+}b$ and $\ell_W(u_H) = \Pi_W(u_H) = 1{+}b$, so $\ell_S(u_H) = \ell_W(u_H)$. This implies, because the intercept of ℓ_S is positive and the intercept of ℓ_W is zero, and the slopes of both lines are positive, the slope of ℓ_W must be greater than that the slope of ℓ_W, i.e., $\beta_W > \beta_S$. Thus, for all $x \in \text{Supp}$, $x \neq u_H$ $\ell_S(x) > \ell_W(x)$. Next note that for all $x \in \text{Supp}$, Remark 2.3.3 and the definition of the reward functions imply that

$$\ell_W(x) = \beta_W\, x = \Pi_W(x) = F_S(x) + \mathbb{1}_r(x),$$
$$\ell_S(x) = \alpha_S + \beta_S\, x = \Pi_S(x) = F_W(x) + \mathbb{1}_r(x).$$

Thus,

$$F_S(x) < F_W(x), \quad x \in \text{Supp}\backslash\{u_H\} \text{ and } F_S(u_H) = F_W(u_L) = 1,$$

i.e., F_S strictly first-order stochastically dominates F_W. Thus, at the upper bond of the rank competition region, u_L, $\Pi_W(u_L) < \Pi_S(u_L)$. For $x \in (u_L, r)$, Π_S and Π_W are constant. At r, the reward functions must jump up and each must meet their respective support lines in order for chasing the bonus to be a best response. The size of the jump will depend on the level of bonus compensation, b, and on mass placed on r. The larger the mass, the larger the jump from topping r. Since the slope of ℓ_W is greater than the slope of ℓ_S, Π_W is lower than Π_S for $x \in (u_L, r)$, and the component of the jump produced by the bonus is the same for both contestants, and the jump must be larger for W, S must place more mass on r. However, both contestants cannot place mass on r, as this would result in a tie with positive probability (Remark 2.3.1). Hence, the mass placed by W must be 0, the mass placed by S, which we denote by p_S^r, must be positive. By the same argument used in the proof of Lemma 2.3.3, W must place positive mass on 0, both contestants, conditioned on choosing performance levels in $(0, u_L]$, where $u_L < r$

must randomize uniformly; both contestants, conditioned on choosing performance levels in $(r, u_H]$ must randomize uniformly.

Let p_i^h, $i = S, W$, represent the probability that contestant i targets the rank and bonus competition region $(r, u_H]$, i.e., chooses performance levels in this region. The arguments above have shown that the performance distributions of the contestants can be described as follows:

(a) Contestant S targets the rank and bonus competition region $(r, u_H]$ with probability p_S^h and, conditioned on targeting this region, randomizes using a uniform distribution Unif$[r, u_H]$; S puts point mass on r with probability p_S^r and targets the pure rank competition $(0, u_L]$ with probability $1 - p_S^h - p_S^r$ and, conditioned on targeting this region, randomizes using a Unif$[0, u_L]$ distribution;

(b) Contestant W targets the rank and bonus competition region $(r, u_H]$ with probability p_W^h and, conditioned on targeting this region, randomizes using a uniform distribution Unif$[r, u_H]$; W puts point mass on 0 with probability p_W^0 and targets the pure rank competition $(0, u_L]$ with probability $1 - p_W^h - p_W^0$ and, conditioned on targeting this region, randomizes using a Unif$[0, u_L]$ distribution.

For contestant S, her reward at bonus threshold should meet her support line. As α_S is just the probability that contestant W plays zero, we have

$$\alpha_S + \beta_S\, r = p_W^0 + \beta_S\, r = (1 - p_W^r) + b. \tag{B.1.15}$$

Meanwhile the slope over $[0, u_L]$ region should be the same as the slope over $[r, u_H]$ region, which both equal to β_S, i.e.,

$$\beta_S = \frac{1 - p_W^h - p_W^0}{u_L} = \frac{p_W^h}{u_H - r}. \tag{B.1.16}$$

Similarly, for contestant W, reward at r meets his support line, i.e.,

$$\beta_W\, r = (1 - p_S^h) + b. \tag{B.1.17}$$

Meanwhile the slope over $[0, u_L]$ region should be the same as the slope over $[r, u_H]$ region, which both equal to β_W, i.e.,

$$\beta_W = \frac{1 - p_S^h - p_S^r}{u_L} = \frac{p_S^h}{u_H - r}. \tag{B.1.18}$$

Thus, we have

$$\frac{1 - p_W^0 - p_W^h}{1 - p_S^r - p_S^h} = \frac{p_W^h}{p_S^h},$$

which implies that

$$\frac{p_W^h}{p_S^h} = \frac{1 - p_W^0}{1 - p_S^r}.$$

Because both contestants use up their capacities in equilibrium, we have

$$(1 - p_S^h - p_S^r)\,\frac{u_L}{2} + r\,p_S^r + p_S^h\,\frac{r + u_H}{2} = \mu_S, \tag{B.1.19}$$

$$(1 - p_W^h - p_W^0)\,\frac{u_L}{2} + p_W^h\,\frac{r + u_H}{2} = \mu_W. \tag{B.1.20}$$

Combining the six equations (B.1.15), (B.1.16), (B.1.17), (B.1.18), (B.1.19) and (B.1.20), we are able to solve for the six unknowns $\{u_L, u_H, p_S^h, p_W^h, p_S^r, p_W^0\}$. The solution can be expressed as follows,

$$p_S^r = \frac{b\,(\mu_S - \mu_W)}{b\,r + \mu_W}, \qquad p_S^h = \frac{(1+b)\,(u_H - r)}{u_H}, \tag{B.1.21}$$

$$p_W^0 = \frac{(1+b)\,(\mu_S - \mu_W)}{b\,r + \mu_S}, \qquad p_W^h = \frac{b\,(u_H - r)}{r - u_L}, \tag{B.1.22}$$

$$u_H = \frac{2\,(1+b)\,(b\,r + \mu_S)\,(b\,r + \mu_W)^2}{b^2\,(b\,r + \mu_S)^2 + (1+b)^2\,(b\,r + \mu_W)^2}, \tag{B.1.23}$$

$$u_L = r - \frac{2\,b\,(b\,r + \mu_S)^2\,(b\,r + \mu_W)}{b^2\,(b\,r + \mu_S)^2 + (1+b)^2\,(b\,r + \mu_W)^2}. \tag{B.1.24}$$

Finally, from the results above it is not hard to calculate the contest reward functions for both contestants. The reward function for contestant S from submitting performance x is as follows,

$$\Pi_S(x) = \begin{cases} p_W^0 + (1 - p_W^0 - p_W^h)\,\dfrac{x}{u_L} & \text{for } x \le u_L, \\ 1 - p_W^h & \text{for } x \in (u_L, r), \\ b + (1 - p_W^h) + p_W^h\,\dfrac{x - r}{u_H - r} & \text{for } x \in [r, u_H], \\ 1 + b & \text{for } x > u_H. \end{cases} \tag{B.1.25}$$

Correspondingly, the reward function for contestant W is

$$\Pi_W(x) = \begin{cases} (1 - p_S^r - p_S^h)\,\dfrac{x}{u_L} & \text{for } x \le u_L, \\ 1 - p_S^r - p_S^h & \text{for } x \in (u_L, r), \\ b + (1 - p_S^h) + p_S^h\,\dfrac{x - r}{u_H - r} & \text{for } x \in [r, u_H], \\ 1 + b & \text{for } x > u_H. \end{cases} \tag{B.1.26}$$

Moreover, the parameters of support lines satisfy

$$\alpha_S = p_W^0, \quad \beta_S = \frac{1 - p_W^h - p_W^0}{u_L}, \quad \beta_W = \frac{1 - p_S^h - p_S^r}{u_L}.$$

Proof of Proposition 2.3.2 Let $p_W^0(\text{Eq2})$ represent the probability that the weak contestant places point mass on zero in an Eq2 equilibrium. Let $p_W^0(\text{Eq0})$ represent the probability that the weak contestant places point mass on zero in an Eq0 equilibrium. Lemma 2.3.1 without bonus shows that $p_W^0(\text{Eq2}) = (\mu_S - \mu_W)/\mu_S$. Substituting the definitions of u_L, u_H and p_W^0 provided in Lemma 2.3.4 and equations (B.1.24) and (B.1.23) in this appendix shows that

$$p_W^0(\text{Eq2}) = p_W^0(\text{Eq0}) \left(1 - \frac{b\,(r - \mu_S)}{b\,r + \mu_S}\right). \tag{B.1.27}$$

It is always the case that

$$0 < \frac{b\,(r - \mu_S)}{b\,r + \mu_S} < 1.$$

Thus, $p_W^0(\text{Eq2}) < p_W^0(\text{Eq0})$.

Proof of Lemma 2.3.5

The proof of this lemma is tedious and so we have broken the steps into a series of results. With this Lemma in hand, the proof of the characterizations in the next proposition, Proposition 2.3.3, is fairly straightforward.

To initiate the proof, for fixed contestant capacities, μ_S and μ_W, define the following functions:

$$\text{ConEq0}(r, b) = \left(1 - \frac{\mu_W}{\mu_S}\right) + \frac{\mu_W}{\mu_S}\,\frac{1}{2\,\mu_S}\,r - (1 + b), \tag{B.1.28}$$

$$\text{ConEq1}(r, b) = \frac{2\,r\,(r - \mu_S)}{(2\,r - u)\,u} - (1 + b), \text{ where } u = \frac{2\,r\,\mu_W}{\mu_W + \sqrt{2\,b\,r\,\mu_W + \mu_W^2}}, \tag{B.1.29}$$

$$\text{ConEq2}(r, b) = u_H - r, \text{ where } u_H = \frac{2\,(1 + b)\,(b\,r + \mu_S)\,(b\,r + \mu_W)^2}{b^2\,(b\,r + \mu_S)^2 + (1 + b)^2\,(b\,r + \mu_W)^2}. \tag{B.1.30}$$

Result B.1.1 *For any bonus compensation package $(r, b) \in (\mu_S, \infty) \times (0, \infty)$,*

(i) An Eq0 equilibrium exists only if ConEq0$(r, b) \geq 0$.
(ii) An Eq1 equilibrium exists only if ConEq1$(r, b) \geq 0$.
(iii) An Eq2 equilibrium exists only if ConEq2$(r, b) > 0$.

Proof: This result is fairly obvious. In an Eq0 equilibrium, $\alpha_S = 1 - \frac{\mu_W}{\mu_S}$, and $\beta_S = \frac{\mu_W}{\mu_S}\frac{1}{2\,\mu_S}$. Thus ConEq0$(r, b) \geq 0$ is equivalent to not chasing the bonus being a best response for S.

As shown in the proof of Lemma 2.3.3, in an Eq1 equilibrium, $\alpha_W = 0$ and $\beta_W = (1-p_S^r)/u$, where u and p_S^r are defined in Lemma 2.3.3. Substituting, the definition of p_S^r also provided in that Lemma, shows that ConEq1 ≥ 0 is equivalent to the condition that $\beta_W r \geq (1+b)$, the condition for not chasing the bonus to be a best reply for W.

By definition, in an Eq2 equilibrium, both contestants pursue rank competition at performance levels in excess of the reward threshold, r. Thus, the expression for the upper bound of the rank and bonus competition region, u_H, defined in equation (B.1.23), must exceed r. □

Result B.1.2 *For any bonus compensation package* $(r, b) \in (\mu_S, \infty)\times(0, \infty)$*,* ConEq0$(r, b) \geq 0$ *if and only if an Eq0 equilibrium exists.*

Proof: As stated above, ConEq0$(r, b) \geq 0$ implies that not chasing the bonus is a best reply for S, ConEq0$(r, b) \geq 0$ implies that $r > 2\,\mu_S$, the upper bound of performance in an Eq0 equilibrium. Thus, at $x = 2\,\mu_S$, the reward to both contestants is 1. Because the support lines cross at $x = 2\,\mu_S$ and because $\alpha_S > 0$ and $\alpha_W = 0$, for $x \geq 2\,\mu_S$, $\beta_W\,x \geq \alpha_S + \beta_S\,x$. Thus, $\alpha_S + \beta_S\,r \geq 1 + b$ implies that $\beta_W\,r \geq 1 + b$. Because, ConEq0$(r, b) \geq 0$ is equivalent to $\alpha_S + \beta_S\,r \geq 1 + b$, ConEq0$(r, b) \geq 0$ also implies that not chasing the bonus is a best reply for W. □

Result B.1.3 *For any bonus compensation package* $(r, b) \in (\mu_S, \infty) \times (0, \infty)$*, an Eq1 equilibrium exists if and only if* ConEq1$(r, b) \geq 0$ *and* ConEq0$(r, b) < 0$*.*

Proof: For an Eq1 equilibrium to exist it must be the case that $p_W^0 \in (0, 1)$, $p_S^r \in (0, 1]$, and $u < r$, where p_W^0, p_S^r and u are defined in Lemma 2.3.3. Using the definition of p_S^r provided in Lemma 2.3.3, and the assumption that $r > \mu_S$, we see that $p_S^r < 1$. Inspecting the definition of p_W^0 provided in Lemma 2.3.3 shows that $p_W^0 < 1$. Using the definitions in Lemma 2.3.3, we see that

$$u = r\,\frac{2\,\mu_W}{\mu_W + \sqrt{2\,b\,r\,\mu_W + \mu_W^2}} \text{ and } \frac{2\,\mu_W}{\mu_W + \sqrt{2\,b\,r\,\mu_W + \mu_W^2}} < 1.$$

Hence, $u < r$. Inspection of the definition of u provided by Lemma 2.3.3 shows that $u > 0$.

Now consider p_W^0. Algebraic simplification shows that ConEq1 can be expressed at follows:

$$\begin{aligned}\text{ConEq1}(r,b) &= \mathcal{S}(r,b)\,\frac{\mu_W\ (b\,r+\mu_W)}{r\,\mu_W\,\sqrt{\mu_W\ (2\,b\,r+\mu_W)}},\ \text{where}\\ \mathcal{S}(r,b) &= (r-\mu_S) - \frac{b\,r+\mu_S}{b\,r+\mu_W}\,\sqrt{\mu_W\ (2\,b\,r+\mu_W)}.\end{aligned} \tag{B.1.31}$$

Thus,

$$\operatorname{sgn}\big(\text{ConEq1}(r,b)\big) = \operatorname{sgn}\big(\mathcal{S}(r,b)\big). \tag{B.1.32}$$

Thus, if ConEq1 ≥ 0, then using the definition of u in Lemma 2.3.3, and equations (B.1.31) and (B.1.32), we see that

$$\mathcal{S}(r,b) \geq 0 \Rightarrow u \geq \frac{2\,r\,\mu_W}{\mu_W + (r-\mu_S)\ \frac{b\,r+\mu_W}{b\,r+\mu_S}}.$$

The right-hand side of this expression is decreasing in b and

$$\lim_{b\to\infty} \frac{2\,r\,\mu_W}{\mu_W + (r-\mu_S)\ \frac{b\,r+\mu_W}{b\,r+\mu_S}} = 2\,\mu_W\,\frac{r}{r-(\mu_S-\mu_W)} > 2\,\mu_W.$$

Thus, $u > 2\,\mu_W$. This implies, using the definition of p_W^0 provided in Lemma 2.3.3, that $p_W^0 > 0$.

Now consider p_S^r. First note that because $u < r$, $p_S^r > 0$ if and only if $2\,\mu_S > u$. Using the quadratic formula, and the definition of u provided in Lemma 2.3.3 we see that

$$\text{ConEq0} < 0 \iff \mu_S > \mu_S^o = \frac{\sqrt{2\,b\,r\,\mu_W + \mu_W^2} - \mu_W}{2\,b}. \tag{B.1.33}$$

Next note, given the definition of u in Lemma 2.3.3,

$$2\,\mu_S - u = 2\,\mu_S - \frac{2\,r\,\mu_W}{\mu_W + \sqrt{2\,b\,r\,\mu_W + \mu_W^2}}.$$

At $\mu_S = \mu_S^o$,

$$2\,\mu_S - \frac{2\,r\,\mu_W}{\mu_W + \sqrt{2\,b\,r\,\mu_W + \mu_W^2}} = 0. \tag{B.1.34}$$

Because, the left hand side of equation (B.1.34) is increasing in μ_S, equations (B.1.33) and (B.1.34) imply that

$$p_S^r > 0 \iff 2\,\mu_S - u > 0 \iff \mu_S > \mu_S^o \iff \text{ConEq0} < 0.$$

□

Result B.1.4 *For any bonus compensation package* $(r, b) \in (\mu_S, \infty) \times (0, \infty)$*, an Eq2 equilibrium exists if and only if ConEq2*$(r, b) > 0$.

Proof: An Eq2 equilibrium exists if and only if all the parameters specified in Lemma 2.3.4 satisfy their relevant range restrictions. First note that u_L, as defined in the Proof of Lemma 2.3.4 by equation (B.1.24), can be shown by algebraic manipulation to be equal to

$$\frac{b^2\, r\, (r-\mu_S)\, (r+2\,b\,r+\mu_S)+2\,b\, (r-\mu_S)\, (r+2\,b\,r+\mu_S)\, \mu_W+(1+b)^2\, r\, \mu_W^2}{b^2\, \left((1+2\,b\,(1+b))\, r^2+2\,b\, r\, \mu_S+\mu_S^2\right)+2\,b\,(1+b)^2\, r\, \mu_W+(1+b)^2\, \mu_W^2} > 0.$$

Thus, $u_L > 0$. Inspecting the definition of u_L provided by equation (B.1.24) shows that $u_L < r$. Now consider p_W^0. The definition of p_W^0 provided in Lemma 2.3.4 states that

$$p_W^0 = \frac{(1+b)\, (\mu_S-\mu_W)}{b\, r+\mu_S},$$

and

$$0 < \frac{(1+b)\, (\mu_S-\mu_W)}{b\, r+\mu_S} < \frac{(1+b)\, (\mu_S-\mu_W)}{b\, \mu_S+\mu_S} = \frac{\mu_S-\mu_W}{\mu_S} < 1.$$

So, $p_W^0 \in (0, 1)$. Now consider p_S^r. The definition of p_S^r provided in Lemma 2.3.4 states that

$$p_S^r = \frac{b\, (\mu_S-\mu_W)}{b\, r+\mu_W},$$

and

$$0 < \frac{b\, (\mu_S-\mu_W)}{b\, r+\mu_W} < \frac{b\, (\mu_S-\mu_W)}{b\, \mu_S+\mu_W} < 1.$$

Now suppose that $p_S^h \in (0, 1)$. The definitions p_S^h and p_W^h provided by equations (B.1.21) and (B.1.22) imply that

$$p_W^h = \frac{b\, r+\mu_W}{b\, r+\mu_S}\, p_S^h.$$

Thus if $p_S^h \in (0, 1)$ then $p_W^h \in (0, 1)$. So an Eq2 equilibrium will exist if $u_H > r$ and $p_S^h \in (0, 1)$.

First we show that it is always the case that $p_S^h < 1$. Note that the definition of u_H provided by equation (B.1.23) implies that $p_S^h \geq 1$ if and only if

$$\frac{2\, (b\, r+\mu_S)\, (b\, r+\mu_W)^{\,2}}{b^2\, (b\, r+\mu_S)^{\,2}+(1+b)^2\, (b\, r+\mu_W)^{\,2}} \geq \frac{r}{b}. \tag{B.1.35}$$

Condition (B.1.35) can only be satisfied if

$$\begin{aligned}\left(2\ (b\,r+\mu_S)\ (b\,r+\mu_W)^{\,2}\right)\ b-\left(b^2\ (b\,r+\mu_S)^{\,2}+(1+b)^2\ (b\,r+\mu_W)^{\,2}\right)\ r=\\ -\,r\,\mu_W^2-2\,b^3\,r^2\ (r-\mu_W)-2\,b\,\mu_W\ \left(r^2+\mu_W\ (r-\mu_S)\right)-\\ b^2\,r\ \left(r^2+\left(\mu_S^2-\mu_W^2\right)+4\,\mu_W\ (r-\mu_S)\right)>0,\end{aligned}$$

which is clearly impossible. Thus, $p_W^h < 1$.

Finally note that from the definition of p_W^h provided by equation (B.1.21), $p_W^h > 0$ if and only if $u_H > r$. Thus, an Eq1 equilibrium exists if and only if $u_H > r$, i.e., ConEq2 > 0. □

Result B.1.5 *$sgn\big(ConEq2\big) = -sgn\big(ConEq1\big)$.*

Proof: Let

$$\mathcal{S}_2(r,b) = (r-\mu_S)^2 - \mu_W\ (2\,b\,r+\mu_W)\ \left(\frac{b\,r+\mu_S}{b\,r+\mu_W}\right)^2.$$

Algebraic simplification shows that

$$\begin{aligned}&\text{ConEq2}(r,b) = -K\,\mathcal{S}_2(r,b),\ \text{where}\\ &K = \frac{(b\,r+\mu_W)^{\,2}}{r\ (((1+b)^2+b^2)\ (b\,r+\mu_W)^{\,2}+(\mu_S-\mu_W)\ b^2\ ((b\,r+\mu_W)+(b\,r+\mu_S)))}.\end{aligned}$$

$K > 0$, so

$$\text{sgn}\big(\text{ConEq2}\big) = -\text{sgn}\big(\mathcal{S}_2(r,b)\big). \tag{B.1.36}$$

Recall $\mathcal{S}$, defined in equation (B.1.31), and note that

$$\text{sgn}\big(\mathcal{S}_2(r,b)\big) = \text{sgn}\big(\mathcal{S}(r,b)\big). \tag{B.1.37}$$

The result follows from (B.1.36), (B.1.37), and (B.1.32). □

Result B.1.6 *Define the parametric curves, with parameter $y > 0$ as follows:*

$$\begin{aligned}\mathcal{B}_0^1(y) &= \frac{\sqrt{2\,y\,\mu_W+\mu_W^2}-\mu_W}{2\,\mu_S}, & \mathcal{R}_0^1(y) &= \frac{2\,y\,\mu_S}{\sqrt{2\,y\,\mu_W+\mu_W^2}-\mu_W};\\ \mathcal{B}_1^2(y) &= y\left(\mu_S+\frac{y+\mu_S}{y+\mu_W}\sqrt{2\,y\,\mu_W+\mu_W^2}\right)^{-1}, & \mathcal{R}_1^2(y) &= \mu_S+\frac{y+\mu_S}{y+\mu_W}\sqrt{2\,y\,\mu_W+\mu_W^2}.\end{aligned}$$

These parametric curves have the following properties:

(i) $\mathcal{B}_0^1(y)$ *and* $\mathcal{B}_1^2(y)$ *are strictly increasing and map* $(0,\infty)$ *onto* $(0,\infty)$.
(ii) $\mathcal{B}_0^1(y)\,\mathcal{R}_0^1(y) = \mathcal{B}_1^2(y)\,\mathcal{R}_1^2(y) = y$.
(iii) $\mathcal{B}_1^2(y) > \mathcal{B}_0^1(y)$.
(iv) $ConEq2(r,b) = 0 \iff (r,b) = (\mathcal{R}_1^2(y), \mathcal{B}_1^2(y))$ *for some* $y > 0$.
(v) $ConEq1(r,b) = 0 \iff (r,b) = (\mathcal{R}_1^2(y), \mathcal{B}_1^2(y))$ *for some* $y > 0$.
(vi) $ConEq0(r,b) = 0 \iff (r,b) = (\mathcal{R}_0^1(y), \mathcal{B}_0^1(y))$ *for some* $y > 0$.

Proof: Property (i) follows from differentiation. Property (ii) follows from inspection. To establish property (iii), first note that $\lim_{y\to 0} \mathcal{B}_1^2(y) - \mathcal{B}_0^1(y) = 0$ and $\lim_{y\to 0}(\mathcal{B}_1^2(y) - \mathcal{B}_0^1(y))' = 0$. Next note that differentiation shows that

$$\begin{aligned}\operatorname{sgn}\big((\mathcal{B}_1^2(y) - \mathcal{B}_0^1(y))'\big) = \\ \operatorname{sgn}\Big(y\ (\mu_S - \mu_W)\ \Big(2\,y^2\,\mu_W + 2\,\mu_S\,\mu_W\ \Big(\mu_W + \sqrt{\mu_W\ (2\,y + \mu_W)}\Big) + \\ y\ \Big(5\,\mu_S\,\mu_W + \mu_W^2 + 2\,\mu_S\,\sqrt{\mu_W\ (2\,y + \mu_W)}\Big)\Big)\Big). \quad \text{(B.1.38)}\end{aligned}$$

Because, $\lim_{y\to 0}(\mathcal{B}_1^2(y) - \mathcal{B}_0^1(y))' = 0$ and $\lim_{y\to 0} \mathcal{B}_1^2(y) - \mathcal{B}_0^1(y) = 0$, to establish the result, it is sufficient to show that the derivative of the right-hand side of (B.1.38) is positive. This follows from differentiation.

To establish parts (iv), (v), first note that, by equations (B.1.32), (B.1.36), and (B.1.37), parts (iv), (v) are equivalent to the assertion that

A1: $$\mathcal{S}(r,b) = 0 \iff \text{ there exists } y > 0 \text{ such that } (r,b) = (\mathcal{R}_1^2(y), \mathcal{B}_1^2(y)).$$

Inspecting the definition of $\mathcal{S}$ provided by equation (B.1.31), we see that $\mathcal{S}(r,b) = 0$ is equivalent to the assertion that

A2: $$\text{there exists } y > 0 \text{ such that } r = \mu_S + \frac{y + \mu_S}{y + \mu_W}\,\sqrt{\mu_W\ (2\,y + \mu_W)} \text{ and } y = r\,b.$$

A1 and A2 are clearly equivalent.

Property (vi) follows from substitution of the definitions of $\mathcal{R}_0^1$ and $\mathcal{B}_0^1$, provided by Result B.1.6, into ConEq0 (equation (B.1.28)). □

Result B.1.7 *If* $(r_1, b) = (\mathcal{R}_0^1(y_1), \mathcal{B}_0^1(y_1))$ *and* $(r_2, b) = (\mathcal{R}_1^2(y_2), \mathcal{B}_1^2(y_2))$, *then* $r_1 > r_2$.

Proof: By parts (i) and (iii) of Result B.1.6, if $b = \mathcal{B}_0^1(y_1) = \mathcal{B}_1^2(y_2)$, then $y_1 > y_2$. By part (ii) of Result B.1.6, $r_1\, b = y_1$ and $r_2\, b = y_2$. Thus $r_1 > r_2$. □

Result B.1.8 *For any fixed $b > 0$,*

(i) $\lim_{r\to\mu} ConEq0(r, b) < 0$ *and* $\lim_{r\to\infty} ConEq0(r, b) > 0$.
(ii) $\lim_{r\to\mu} ConEq1(r, b) < 0$ *and* $\lim_{r\to\infty} ConEq1(r, b) > 0$.
(iii) $\lim_{r\to\mu} ConEq2(r, b) > 0$ *and* $\lim_{r\to\infty} ConEq2(r, b) < 0$.

Proof: These assertions follow from straightforward calculations. □

Result B.1.9 *For each $b > 0$,*

(i) there exits a unique $r_2 > \mu_S$, such that $ConEq1(r_2, b) = ConEq2(r_2, b) = 0$; if $r < r_2$, $ConEq1(r_2, b) < 0$ and $ConEq2(r_2, b) > 0$; if $r > r_2$, $ConEq1(r_2, b) > 0$ and $ConEq2(r_2, b) < 0$.
(ii) There exists a unique $r_1 > \mu_S$, such that $ConEq0(r_1, b) = 0$; if $r < r_1$, $ConEq0(r_1, b) < 0$; if $r > r_1$, $ConEq0(r_1, b) > 0$.
(iii) $r_1 > r_2$.

Proof: The proofs of parts (i) and (ii) are identical. So we will only present the proof of (i). Suppose that $\text{ConEq1}(r_2', b) = 0$ and $\text{ConEq1}(r_2'', b) = 0$. Then, by parts (v) and (iv) of Result B.1.6, $(r_2', b) = (\mathcal{R}_1^2(y'), \mathcal{B}_1^2(y'))$ for some $y' > 0$ and $(r_2'', b) = (\mathcal{R}_1^2(y''), \mathcal{B}_1^2(y''))$ for some $y'' > 0$. By part (i) of Result B.1.6, $\mathcal{B}_1^2(y') = \mathcal{B}_1^2(y'')$ implies that $y' = y''$ and thus $r_2' = r_2''$.

The fact that there is a unique $r = r_2$, such that $\text{ConEq1}(r_2, b) = 0$, combined with the continuity of ConEq1 implies that the sign of ConEq1 is constant for $r < r_2$ and is constant for $r > r_2$. Part (i) thus follows from part (ii) of Result B.1.9 and Result B.1.5.

Part (iii) follows from parts (v), (iv), (vi) of Result B.1.6, and Result B.1.7. □

Result B.1.10 *Define*

$$E = (\mu_S, \infty) \times (0, \infty),$$
$$E_{01} = \{(r, b) : \exists y > 0 \text{ such that } r \geq \mathcal{R}_0^1(y) \text{ and } b = \mathcal{B}_0^1(y)\}.$$
$$E_{12} = \{(r, b) : \exists y > 0 \text{ such that } r \geq \mathcal{R}_1^2(y) \text{ and } b = \mathcal{B}_1^2(y)\}.$$

Then

(i) $(r, b) \in E_{01} \iff ConEq0(r, b) \geq 0$,
(ii) $(r, b) \in E_{12} \iff ConEq1(r, b) \geq 0$,
(iii) $(r, b) \in E \backslash E_{12} \iff ConEq2(r, b) > 0$.

Proof: The proofs of parts (i), (ii) are virtually identical. Thus we will only prove part (i). We first prove sufficiency. Note that if $(r, b) = (\mathcal{R}_0^1(y), \mathcal{B}_0^1(y))$, then by part (vi) of Result B.1.6, ConEq0$(r, b) = 0$. Next suppose that $b = \mathcal{R}_0^1(y)$ and $r > \mathcal{R}_0^1(y)$. Let $r_1 = \mathcal{R}_0^1(y)$. Part (ii) of Result B.1.9 shows that r_1 is the unique r satisfying ConEq0$(r, b) = 0$, and implies, because $r > r_1$, ConEq0$(r, b) > 0$.

To prove necessity, suppose that $(r, b) \notin E_{01}$. Let y' be the unique $y > 0$ such that $\mathcal{B}_0^1(y) = b$ (the existence and uniqueness of y' follows from property (i) in Result B.1.6). Thus, $\mathcal{B}_0^1(y') = b$. If $(r, b) \notin E_{01}$, then it must be the case that $r < \mathcal{R}_0^1(y)$, using the same argument as used above for sufficiency, we see that Part (ii) of Result B.1.9 implies that ConEq0$(r, b) < 0$.

Part (iii) follows from part (ii) and Result B.1.5. □

Proof of Lemma 2.3.5 Result B.1.10, and Results B.1.2, B.1.3, and B.1.4 show that an Eq0 equilibrium can be sustained if and only if $(r, b) \in E_{01}$; an Eq1 equilibrium can be sustained if and only if $(r, b) \in E_{12} \backslash E_{01}$; an Eq2 equilibrium can be sustained if and only if $(r, b) \in E \backslash E_{12}$. We need only show that these three sets are disjoint. Clearly E_{01} and $E_{12} \backslash E_{01}$ are disjoint as are $E_{12} \backslash E_{01}$ and $E \backslash E_{12}$. Now consider E_{01} and $E \backslash E_{12}$. Result B.1.9 implies that $E_{01} \subset E_{12}$. Thus E_{01} and $E \backslash E_{12}$ are disjoint.

Result B.1.11 *For $r_o > \mu_S$, $r_o > \inf_{y>0} \mathcal{R}_1^2(y)$ if and only if for some $b > 0$ there exists $r < r_o$ such that (r, b) sustains an Eq1 equilibrium.*

Proof: We start by proving sufficiency. Suppose that $r_o > \inf_{y>0} \mathcal{R}_1^2(y)$, then there

exists some $y > 0$, say y_2, such that $r_o > \mathcal{R}_1^2(y_2)$. Let $(r_2, b_o) = (\mathcal{R}_1^2(y_2), \mathcal{B}_1^2(y_2))$. By property (v) in Lemma B.1.6, $\text{ConEq1}(r_2, b_o) = 0$. The fact that $r_o > r_2$, implies, by part (i) of Result B.1.9, that $\text{ConEq1}(r_o, b_o) > 0$.

Parts (ii) and (iii) of Result B.1.9 show that there is a unique bonus threshold, r_1, such that $\text{ConEq0}(r_1, b_o) = 0$ and that $r_1 > r_2$. Let $r^* = \min[r_1, r_o]$. Because $r_1 > r_2$ and $r_o > r_2$, $r^* > r_2$. By definition, $r^* \leq r_o$. Hence, if $r \in (r_2, r^*)$, $r < r_o$, $r > r_1$ and $r < r_2$. Parts (ii) and (i) of Result B.1.9 and Result B.1.3 thus establish that (r, b) sustains an Eq1 equilibrium.

To prove necessity, note simply that $r_o \leq \inf_{y>0} \mathcal{R}_1^2(y)$ implies that the set, $\{(r, b) : r < r^o\} \cap E_{12}$ is empty. And thus, by part (ii) of Result B.1.10, for all $b > 0$, if $r < r^o$ then $\text{ConEq1}(r_2, b_o) < 0$, and thus, by Result B.1.3, an Eq1 equilibrium cannot be sustained. □

Proof of Proposition 2.3.3

Proof of Parts (i) **and** (ii) **of Proposition 2.3.3** By definition provided in Result B.1.6, $\mathcal{R}_1^2(y') = 2\,\mu_S$, is equivalent to

$$2\,\mu_S = \mu_S + \frac{y + \mu_S}{y + \mu_W}\sqrt{2\,y\,\mu_W + \mu_W^2}.$$

The equation $\frac{y+\mu_S}{y+\mu_W}\sqrt{2\,y\,\mu_W + \mu_W^2} - \mu_S = 0$ has solution with $y > 0$ if and only if $y \hookrightarrow (\frac{y+\mu_S}{y+\mu_W})^2\,(2\,y\,\mu_W + \mu_W^2) - \mu_S^2$ has a positive real root. After algebraic simplification we see that

$$\left(\frac{y + \mu_S}{y + \mu_W}\right)^2 (2\,y\,\mu_W + \mu_W^2) - \mu_S^2 = \frac{y}{(y + \mu_W)^2}\,\text{Poly}(y), \quad \text{where}$$
$$\text{Poly}(y) = 2\,\mu_W\,y^2 + \mu_W^2\left(\left(\sqrt{5} - 2\right) + \frac{\mu_S}{\mu_W}\right)\left(\left(2 + \sqrt{5}\right) - \frac{\mu_S}{\mu_W}\right) y + 2\,\mu_S\,\mu_W^2.$$

Taking the discriminant of Poly we obtain

$$\text{Disc} = (\mu_S - \mu_W)^2\,\mu_W\,\mu_S\left(\frac{\mu_S}{\mu_W} + \frac{\mu_W}{\mu_S} - 6\right). \tag{B.1.39}$$

Let $a = {}^{\mu_S}/_{\mu_W} > 1$. Disc ≤ 0 is equivalent to $a + {}^1/_a - 6 \leq 0$. Thus, if $a \leq 3 + 2\sqrt{2}$, Disc ≤ 0 and thus Poly(y) either has no roots or a double root at 0. Hence, $\mathcal{R}_1^2(y) - 2\,\mu_S \geq 0$, for all $y > 0$. Thus, $\inf_{y>0} \mathcal{R}_1^2(y) \geq 2\,\mu_S$. Thus, part (i) follows from Result B.1.11.

If $a > 3 + 2\sqrt{2}$ then Poly(y) has two real roots. Because Poly is convex and is positive at $y = 0$, both roots are either positive or negative. Because $3 + 2\sqrt{2} > 2 + \sqrt{5}$, $a > 3 + 2\sqrt{2}$ implies that $\left(2 + \sqrt{5}\right) - a < 0$. This implies that Poly$'(0) < 0$ which, in turn, implies that both roots are positive. Hence, there exists a y-interval over which Poly$(y) < 0$. This implies that, over this interval, $\mathcal{R}_1^2(y) - 2\,\mu_S < 0$. Hence, $\inf_{y>0} \mathcal{R}_1^2(y) < 2\,\mu_S$. Hence, part (ii) follows from Result B.1.11.

Proof of part (iii) **of Proposition 2.3.3** First note that the upper bound of the rank plus bonus competition region in the Eq2 configuration, u_L provided by equation (B.1.23) in the proof of Lemma 2.3.4 is given by

$$u_H(b) = \frac{2\ (b\,r + \mu_S)\ (b\,r + \mu_W)^2\ (1 + b)}{b^2\ (b\,r + \mu_S)^2 + (1 + b)^2\ (b\,r + \mu_W)^2}. \tag{B.1.40}$$

Next note that

$$\lim_{b \to \infty} u_H(b) = r.$$

Differentiating u_H and evaluating the derivative at 0 yields

$$u_H'(0) = 2\,(r - \mu_S) > 0. \tag{B.1.41}$$

Thus u_H is increasing for b sufficiently small. To verify that u_H is asymptotically decreasing, define the function

$$u_H^o(q) = u_H(1/q), \quad q > 0. \tag{B.1.42}$$

u_H will be asymptotically decreasing if and only if u_H^o is increasing in a neighborhood of 0. First note that

$$u_H^o{}'(q) = \frac{2\,(1 + q)\ (r + q\,\mu_S)\ (r + q\,\mu_W)^{\,2}}{(2 + q\,(2 + q))\,r^2 + 2\,q\,r\ (\mu_S + (1 + q)^2\,\mu_W) + q^2\ (\mu_S^2 + (1 + q)^2\,\mu_W^2)}. \tag{B.1.43}$$

Taking the limit of $u_H^o{}'$ as $q \to 0$ yields

$$\lim_{q \to 0} u_H^o{}'(q) = \mu_W > 0. \tag{B.1.44}$$

Thus, u_H^o is increasing in a neighborhood of 0 and thus u_H is asymptotically decreasing.

To establish the bounds on u_H asserted in the theorem, first note from the proof of Proposition that $u_H \geq r$, were u_H is defined by equation (B.1.23), is a necessary and sufficient condition for the existence of an Eq2 equilibrium. This verifies the lower bound.

Using the definition of u_H provided by equation (B.1.23), and making the substitution,

$y = b\,r$ yields,

$$\frac{r}{u_H(y)} = \frac{1}{2}\left(\left(\frac{y}{r+y}\,\frac{y}{\mu_W+y}\,\frac{\mu_S+y}{\mu_W+y}\right) + \frac{r+y}{\mu_S+y}\right). \tag{B.1.45}$$

The Arithmetic/Geometric mean inequality implies that

$$\frac{r}{u_H(y)} \geq \sqrt{\left(\frac{y}{r+y}\,\frac{y}{\mu_W+y}\,\frac{\mu_S+y}{\mu_W+y}\right)\frac{r+y}{\mu_S+y}} = \frac{y}{\mu_W+y}. \tag{B.1.46}$$

Using the fact that $y = r\,b$ and simplifying shows that

$$u_H \leq r + \frac{\mu_W}{b}. \tag{B.1.47}$$

Next note that we can also express r/u_H as

$$\frac{r}{u_H(y)} = \frac{y}{\mu_W+y}\,\frac{1}{2}\left(\frac{\mu_S+y}{r+y}\,\frac{y}{\mu_W+y} + \left(\frac{\mu_S+y}{r+y}\,\frac{y}{\mu_W+y}\right)^{-1}\right). \tag{B.1.48}$$

Let

$$z = \frac{\mu_S+y}{r+y}\,\frac{y}{\mu_W+y}. \tag{B.1.49}$$

Note that the function $y \to \frac{\mu_S+y}{r+y}\,\frac{y}{\mu_W+y}$ is strictly increasing and that the image of $[0,\infty]$ under this function is $[0,1]$. Its inverse function is given by,

$$y = \frac{z\,(r+\mu_W) - \mu_S + \sqrt{4\,r\,(1-z)\,z\,\mu_W + (z\,(r+\mu_W) - \mu_S)^2}}{2\,(1-z)}. \tag{B.1.50}$$

Using, equations (B.1.49) and (B.1.50) we can express (B.1.48) in terms of z as follows:

$$\frac{r}{u_H(z)} = \frac{\text{Num}(z)}{\text{Denom}(z)} = \frac{r\,(1+z^2)}{z\,(r-\mu_W) + \mu_S + \sqrt{4\,r\,(1-z)\,z\,\mu_W + (\mu_S - z\,(r+\mu_W))^2}},\ z \in [0,1]. \tag{B.1.51}$$

Because $\text{Denom}'' \geq 0$, Denom is convex and thus

$$\begin{aligned}\text{Denom}(z) = \text{Denom}((1-z)\,0 + z\,1) \leq (1-z)\,\text{Denom}(0) + z\,\text{Denom}(0) = \\ 2\,((1-z)\,\mu_S + z\,r),\ z \in [0,1].\end{aligned} \tag{B.1.52}$$

Equations (B.1.51) and (B.1.52) imply that

$$\frac{r}{u_H(z)} \geq \frac{r\,(1+z^2)}{2\,((1-z)\,\mu_S + z\,r)},\ z \in [0,1]. \tag{B.1.53}$$

Minimizing the right-hand side of the equation above shows that

$$\frac{r\,(1+z^2)}{2\,((1-z)\,\mu_S+z\,r)} \geq \min_{z\in[0,1]} \frac{r\,(1+z^2)}{2\,((1-z)\,\mu_S+z\,r)} = \frac{r}{\mu_S+\sqrt{(r-\mu_S)^2+\mu_S^2}},\ z\in[0,1]. \tag{B.1.54}$$

Equations (B.1.53) and (B.1.54) imply that

$$\frac{r}{u_H(z)} \geq \frac{r}{\mu_S+\sqrt{(r-\mu_S)^2+\mu_S^2}}. \tag{B.1.55}$$

Thus,

$$u_H \leq \mu_S+\sqrt{(r-\mu_S)^2+\mu_S^2}. \tag{B.1.56}$$

Combining (B.1.47) and (B.1.56) yields the result.

Condition for u_H crossing once the upper bound of pure rank competition risk taking, Lemma B.1.1

Lemma B.1.1 *Suppose that strength asymmetry between the two contestants is not extreme, i.e., $\mu_S/\mu_W < 3+2\sqrt{2}$ and that $r < 2\,\mu_S$. Define,*

$$\lambda^o = \frac{1}{6}\left(-\sqrt{2\left(18-7\sqrt{3}\right)}-3\sqrt{3}+9\right) \approx 0.063,\ \rho^o = \frac{1}{2}\left(\sqrt{6\left(18-7\sqrt{3}\right)}-3\sqrt{3}+9\right) \approx 4.87.$$

If either of the following conditions is satisfied:

(i) $\mu_S/\mu_W < \rho^o$, *or*
(ii) $r < (2-\lambda^o)\mu_S$,

then there exists a unique cut off bonus level $0 < b^o < \infty$ such that the upper bound for contestant performance under rank and bonus competition is greater (less) than the upper bound under pure rank competition if and only if $b < (>)\, b^o$.

Proof: The upper bound of performance under pure rank competition is $2\,\mu_S$. We aim to provide conditions under which the upper bound under bonus competition is greater than the upper bound under rank and bonus competition. As shown by Proposition 2.3.3, the hypotheses of the theorem imply that, for all $b > 0$, all equilibria are Eq2 equilibria. Thus, the question we consider reduces to finding conditions under which u_H (defined in

equation (B.1.23)) is greater (less) than $2\,\mu_S$, or equivalently, conditions under which

$$\frac{r}{u_H} < (>) \frac{r}{2\,\mu_S}. \tag{B.1.57}$$

As in the proof of part (iii) of Proposition 2.3.3, we resolve this inequality by using the substitution, $b = y/r$ to replace b in the expression for u_H provided by equation (B.1.23), an then analyze the conditions on y required to verify

$$\frac{r}{u_H(y)} = \frac{1}{2}\left(\frac{y}{r+y}\,\frac{y}{\mu_W+y}\,\frac{\mu_S+y}{\mu_W+y} + \frac{r+y}{\mu_S+y}\right) < (>) \frac{2}{2\mu_S}, \quad y = r\,b. \tag{B.1.58}$$

Algebraic simplification shows that

$$\operatorname{sgn}(\Phi(y)) = \operatorname{sgn}\left(\frac{1}{2}\left(\frac{y}{r+y}\,\frac{y}{\mu_W+y}\,\frac{\mu_S+y}{\mu_W+y} + \frac{r+y}{\mu_S+y}\right) - \frac{r}{2\,\mu_S}\right), \text{ where} \tag{B.1.59}$$

$$\Phi(y) = (y+\mu_S)^2\, y\,\mu_S - (y+\mu_W)^2\,(r+y)\,(r-\mu_S). \tag{B.1.60}$$

Remark B.1.1. Any result that implies that the discriminant of Φ is negative implies the conclusions of the lemma: Because, $\Phi(0) < 0$, and $\lim_{y\to\infty}\Phi = \infty$, the continuity of Φ implies that there exists $y^o > 0$ such that $\Phi(y^o) = 0$. Because Φ is a cubic polynomial, if Φ has a negative discriminant, Φ has one and only one real root (Chapt. 5.5, Theorem 8, Birkhoff and MacLane, 1977). Thus, a negative discriminant implies that there exists a unique $y^o > 0$ such that $\Phi(y) < (>)\,0$ if and only if $y < (>)\,y^o$. Letting $b^o = y^o/r$, this implies that

$$\frac{r}{u_H} < (>)\,\frac{r}{2\,\mu_S} \Longleftrightarrow b < (>)\,b^o,$$

which immediately implies that

$$u_H > (<)\,2\,\mu_S \Longleftrightarrow b < (>)\,b^o.$$

In the remainder of this proof we will use the notation Disc_x to represent the discriminant of a polynomial in x, where x represents a generic variable. Using this notation and performing some manipulations shows that

$$\begin{aligned}\operatorname{sgn}(\text{Disc}_y[\Phi]) = \operatorname{sgn}\Big(&-4\,r^4\,\mu_W + 4\,r^3\,\mu_W\,(\mu_S+3\,\mu_W) - 4\,\mu_S\,\left(\mu_S^2+\mu_W^2\right)^2 + \\ &r^2\,\left(\mu_S^3 + 18\,\mu_S^2\,\mu_W - 39\,\mu_S\,\mu_W^2 - 12\,\mu_W^3\right) + \\ &4\,r\,\mu_W\,\left(-5\,\mu_S^3 + 9\,\mu_S^2\,\mu_W + 3\,\mu_S\,\mu_W^2 + \mu_W^3\right)\Big).\end{aligned} \tag{B.1.61}$$

Next perform the variable substitutions,

$$\begin{aligned} r = (1-\lambda)\,2\,\mu_S + \lambda\,\mu_S \quad \text{and} \quad \mu_S = \rho\mu_W, \\ 0 < \lambda < 1 \quad \text{and} \quad 1 < \rho < 3 + 2\sqrt{2}, \end{aligned} \tag{B.1.62}$$

where the restrictions on the parameter range follow from the hypotheses of the lemma. After performing these substitutions and simplifying we see that the right hand side of equation (B.1.61) has the same sign as the polynomial, $\mathcal{P}$ defined below:

$$\mathcal{P} = \overbrace{\Big((\lambda-4)\,\lambda\Big)}^{a_4}\,\rho^4 + \overbrace{\Big(2\,\lambda\,\left(14 - 27\,\lambda + 14\,\lambda^2 - 2\,\lambda^3\right)\Big)}^{a_3}\,\rho^3 + \overbrace{\Big(4 - 24\,\lambda + 33\,\lambda^2 - 12\,\lambda^3\Big)}^{a_2}\,\rho^2 + \overbrace{\Big(12\,(3\,\lambda - \lambda^2 - 2)\Big)}^{a_1}\,\rho + \overbrace{\Big(4 - 4\,\lambda)\Big)}^{a_0}. \tag{B.1.63}$$

In equation (B.1.63) we expressed $\mathcal{P}$ as a polynomial in ρ. We can also express $\mathcal{P}$ as a polynomial in λ as follows

$$\mathcal{P} = \overbrace{\Big(-4\,\rho^3\Big)}^{b_4}\,\lambda^4 + \overbrace{\Big(4\,\rho^2\,(7\,\rho - 3)\Big)}^{b_3}\,\lambda^3 + \overbrace{\Big(\rho\,(\rho\,(33 + (\rho - 54)\,\rho) - 12)\Big)}^{b_2}\,\lambda^2 + \overbrace{\Big(-4\,(1 + \rho\,(-9 + (\rho-6)\,(\rho-1)\,\rho))\Big)}^{b_1}\,\lambda + \overbrace{\Big(4\,(1 + (\rho-6)\,\rho)\Big)}^{b_0}. \tag{B.1.64}$$

Remark B.1.2. These observations have established that the sign of $\mathrm{Disc}_y(\Phi)$ equals the sign of $\mathcal{P}$.

Next we show that both condition (i) and condition (ii) are sufficient conditions for $\mathcal{P} < 0$. The key to establishing these results is signing the discriminants of $\mathcal{P}$ when $\mathcal{P}$ is viewed as a quartic polynomial in ρ and when it is viewed as a quartic polynomial in λ. These discriminants will provide information about the number real roots, which when combined with other characteristics of the polynomials, will identify the sign of $\mathcal{P}$.

Result B.1.12 *If $\rho < \rho^o$, then* $\mathrm{Disc}_\lambda[\mathcal{P}] > 0$.

Proof: First note that

$$\mathrm{Disc}_\lambda[\mathcal{P}] = 256\,(\rho-1)^2\,\rho^8\,(1+\rho)\,\left(\rho^4 - 18\,\rho^3 + 54\,\rho^2 + 54\,\rho - 27\right)^3. \tag{B.1.65}$$

Thus, because $\rho > 1$,

$$\text{sgn}(\text{Disc}_\lambda[\mathcal{P}]) = \text{sgn}(\rho^4 - 18\,\rho^3 + 54\,\rho^2 + 54\,\rho - 27). \tag{B.1.66}$$

The polynomial on the right hand side of equation (B.1.66) has four real roots, one negative and three positive. These roots are listed in order of magnitude below:

$$\overbrace{\frac{1}{2}\left(9-3\sqrt{3}-\sqrt{6\left(18-7\sqrt{3}\right)}\right)}^{\approx -1.07},\overbrace{\frac{1}{2}\left(9+3\sqrt{3}-\sqrt{6\left(18+7\sqrt{3}\right)}\right)}^{\approx 0.376},\overbrace{\frac{1}{2}\left(9-3\sqrt{3}+\sqrt{6\left(18-7\sqrt{3}\right)}\right)}^{\rho^o},\overbrace{\frac{1}{2}\left(9+3\sqrt{3}+\sqrt{6\left(18+7\sqrt{3}\right)}\right)}^{\approx 13.8}.$$

Because this polynomial is monomial quartic, its limit is infinity as $\rho \to \pm\infty$ is infinity; hence, at the first root the polynomial crosses the x-axis from above, at the second from below, and at its third crossing, ρ^o, it crosses the x-axis from above. Because 1 is greater than the second root, this polynomial is positive between 1 and ρ^o, implying that $\text{Disc}_\lambda[\mathcal{P}] > 0$. □

Result B.1.13 *If $\lambda > \lambda^o$, then $\text{Disc}_\rho[\mathcal{P}] < 0$.*

Proof: First note that

$$\text{Disc}_\rho[\mathcal{P}] = 256\,(2-\lambda)^2\,(1-\lambda)\,\lambda\,\left(27\,\lambda^4 - 162\,\lambda^3 + 270\,\lambda^2 - 144\,\lambda + 8\right)^3. \tag{B.1.67}$$

Thus, because $\lambda \in (0,1)$,

$$\text{sgn}(\text{Disc}_\rho[\mathcal{P}]) = \text{sgn}(27\,\lambda^4 - 162\,\lambda^3 + 270\,\lambda^2 - 144\,\lambda + 8). \tag{B.1.68}$$

The polynomial on the right hand side of equation (B.1.68), as four positive roots:

$$\overbrace{\frac{1}{6}\left(9-3\sqrt{3}-\sqrt{2\left(18-7\sqrt{3}\right)}\right)}^{\lambda^o},\overbrace{\frac{1}{6}\left(9+3\sqrt{3}-\sqrt{2\left(18+7\sqrt{3}\right)}\right)}^{\approx 1.07},\overbrace{\frac{1}{6}\left(9-3\sqrt{3}+\sqrt{2\left(18-7\sqrt{3}\right)}\right)}^{\approx 1.21},\overbrace{\frac{1}{6}\left(9+3\sqrt{3}+\sqrt{2\left(18+7\sqrt{3}\right)}\right)}^{\approx 3.66}.$$

Its smallest root is λ_0. The limit of the polynomial as $\lambda \to \pm\infty$ is infinity, thus, at λ_0 the polynomial crosses the x-axis from above. Because, the smallest root greater than λ^o is greater than 1, this implies for $\lambda \in (\lambda^o, 1)$ the polynomial is negative which implies that $\text{Disc}_\rho[\mathcal{P}] < 0$. □

Result B.1.14 *If $\rho < \rho^o$, $\mathcal{P} < 0$ for all $\lambda \in (0,1)$.*

Proof: Transform the polynomial $\lambda \to \mathcal{P}$ defined by equation (B.1.64) into a monic polynomial by dividing the polynomial by the coefficient on its highest term, b_4. This yields,

$$\lambda^4 + \left(-7 + \frac{3}{\rho}\right) \lambda^3 - \frac{-12 + \rho\,(33 + (-54 + \rho)\,\rho)}{4\,\rho^2}\,\lambda^2 + \left(-7 + \frac{1}{\rho^3} - \frac{9}{\rho^2} + \frac{6}{\rho} + \rho\right) \lambda - \frac{1 + (-6 + \rho)\,\rho}{\rho^3}. \quad \text{(B.1.69)}$$

Next, perform the variable substitution (see Chapt. 5.6, Theorem 9, Birkhoff and MacLane, 1977),

$$\lambda = z - \frac{1}{4}\left(-7 + \frac{3}{\rho}\right).$$

to yield the depressed monic polynomial

$$\begin{aligned} &z^4 + q\,z^2 + r\,z + s, \text{ where} \\ q &= -\frac{3 - 60\,\rho + 39\,\rho^2 + 2\,\rho^3}{8\,\rho^2}, \\ r &= \frac{-1 - 78\,\rho + 96\,\rho^2 - 18\,\rho^3 + \rho^4}{8\,\rho^3}, \\ s &= \frac{-3 + 984\,\rho - 1746\,\rho^2 + 548\,\rho^3 + 221\,\rho^4 + 252\,\rho^5}{256\,\rho^4}. \end{aligned} \quad \text{(B.1.70)}$$

Because the roots of the monic depressed quartic differ from the original quartic only by translation, the sign of the discriminant for the depressed monic polynomial as well as the number of real and complex roots are the same as for $\mathcal{P}$. Thus, Result B.1.12 ensures that $\text{Disc}_z[z^4 + q\,z^2 + r\,z + s] > 0$. This ensures that all roots of depressed monic polynomial are real or all roots are complex. If $s > q^2/4$ all roots of the depressed monic polynomial are complex (Rees, 1922). Next, note that

$$s - \frac{1}{4}\,q^2 = \frac{(\rho - 1)\;(3 - 333\,\rho + 1062\,\rho^2 - 242\,\rho^3 + 23\,\rho^4 - \rho^5)}{64\,\rho^4}. \quad \text{(B.1.71)}$$

Because $\rho > 1$, to show that $s - q^2/4 > 0$ we need only show that

$$p(\rho) = 3 - 333\,\rho + 1062\,\rho^2 - 242\,\rho^3 + 23\,\rho^4 - \rho^5 > 0, \quad \rho \in (1, \rho^o).$$

To show this, first note that at $\rho = 1$ and $\rho = \rho^o$, p is positive. So, if $p < 0$ at some point in $(1, \rho^o)$, it would have to be the case that the minimum of p over the interval is negative and attained at some point, say ρ^*, in $(1, \rho^o)$. Using Euclidean division we can divide p by p' to yield $p(\rho) = p_1(r)p'(\rho) + p_r(\rho)$, where p_r is the remainder after division. Because, $p'(\rho^*) = 0$, it must be the case that $p(\rho^*) = p_r(\rho^*)$. Computing the remainder polynomial,

p_r, shows that

$$p_r(\rho) = \frac{16}{25}\left(-474 + 2637\,\rho - 48\,\rho^2 - 19\,\rho^3\right).$$

Consequently, if $p < 0$, at any point in $(1, \rho^o)$, there must exist $\rho^* \in (1, \rho^o)$, such that $p(\rho^*) = p_r(\rho^*) < 0$. However, in fact,

$$p_r(\rho) = \frac{16}{25}\left(-474 + 2637\,\rho - 48\,\rho^2 - 19\,\rho^3\right) > 0, \quad \rho \in [1, \rho^o].$$

To see this, first note that, because $\rho > 1$, $48\,\rho^2 + 19\,\rho^3 < (48 + 19)\,\rho^3 = 67\rho^3$. Thus,

$$p_r(\rho) = \tfrac{16}{25}\left(-474 + 2637\,\rho - 48\,\rho^2 - 19\,\rho^3\right) > \tfrac{16}{25}\left(-474 + 2637\,\rho - 67\,\rho^3\right),\ \rho \in [1, \rho^o]. \quad \text{(B.1.72)}$$

The right-hand side expression in equation (B.1.72) is concave in ρ. Therefore, the expression's minimum is attained at one of the endpoints, 1 or ρ^o. Evaluating the expression at these endpoints shows that it is positive at both end points. Thus, the expression is positive over $\rho \in (1, \rho^o)$. Equation (B.1.72) shows that p_r is positive over $\rho \in (1, \rho^o)$. This contradicts p being negative at its minimum value and thus being negative at any point $\rho \in (1, \rho^o)$.

Thus, for $\rho \in (1, \rho^o)$, $s - q^2/4 > 0$. Therefore, all roots of $\mathcal{P}$ are complex. Because, by hypothesis, $\rho < \rho^o < 3 + 2\sqrt{2}$, $\mathcal{P}$ is negative at $\lambda = 0$ and, consequently, $\mathcal{P} < 0$ for all $\lambda \in (0, 1)$. □

Result B.1.15 *If $\lambda < \lambda^o$, $\mathcal{P} < 0$ for all $\rho \in (1, 3 + 2\sqrt{2})$.*

Proof: By Result B.1.13, the discriminant of $\rho \to \mathcal{P}(\rho)$ is negative. Because $\mathcal{P}$ is a quartic polynomial, this implies that $\mathcal{P}$ has two real and two complex roots (Rees, 1922). Next note that

$$\mathcal{P}(0) = 4(1 - \lambda) > 0,\ \mathcal{P}(1) = -4\left(2 + (-2 + \lambda)\,\lambda\right)^2 < 0,\ \text{and}\ \lim_{\rho \to -\infty} \mathcal{P}(\rho) = -\infty.$$

Thus $\mathcal{P}$ has two roots less than 1. Because $\mathcal{P}$ has at most two roots, it must have no roots over $[1, 3 + 2\sqrt{2}]$. Thus, because $\mathcal{P}(1) < 0$, it must be the case that $\mathcal{P}(\rho) < 0$, for all $\rho \in [1, 3 + 2\sqrt{2}]$. □

The lemma follows from Remarks B.1.1 and B.1.2, and Results B.1.14 and B.1.15. □

Proof of Proposition 2.3.4 First, note that, as was shown in the proof of Result B.1.3, in all Eq1 equilibria, $p_W^0 > 0$. Thus all such equilibria feature some downside risk taking. As was shown in the proof of Result B.1.4, in all Eq2 equilibria, $p_W^0 > 0$. Thus, all such equilibria feature some downside risk taking. Lemma 2.3.1 shows that downside risk taking occurs in Eq0 equilibria. Thus, downside risk taking is never 0 in any equilibrium configuration.

In order to prove that a sequence of bonus thresholds and bonus payments (r_n, b_n) exist that support equilibria in which p_W^0 is arbitrarily small, we will employ the parametric curve, $y \hookrightarrow (\mathcal{R}_1^2(y), \mathcal{B}_1^2(y))$, defined in Result B.1.6.

Define the sequence of bonus packages (r_n, b_n) as follows:

$$r_n = \mathcal{R}_1^2(n) = \mu_S + \frac{n + \mu_S}{n + \mu_W} \sqrt{\mu_W \ (2\, n + \mu_W)}, \tag{B.1.73}$$

$$b_n = \mathcal{B}_1^2(n) = n \left(\mu_S + \frac{n + \mu_S}{n + \mu_W} \sqrt{\mu_W \ (2\, n + \mu_W)} \right)^{-1}. \tag{B.1.74}$$

Result B.1.6 shows that for all n, (r_n, b_n) sustains an Eq1 equilibrium and only an Eq1 equilibrium. Using the definition of u, the upper bond of the rank plus bonus competition in Eq1 configurations provided in Lemma 2.3.3, we see that

$$u_n = 2\,\mu_W \left(\frac{\mu_S}{\mu_W + \sqrt{\mu_W \ (2\, n + \mu_W)}} + \frac{n + \mu_S}{n + \mu_W} \frac{\sqrt{\mu_W \ (2\, n + \mu_W)}}{\mu_W + \sqrt{\mu_W \ (2\, n + \mu_W)}} \right). \tag{B.1.75}$$

The first term in the parentheses on the right-hand side of equation (B.1.75) converges to 0 as $n \to \infty$. The second term in the parentheses on the right-hand side of equation (B.1.75) converges to 1. Thus, $u_n \to 2\,\mu_W$. The definition of downside risk taking, p_W^0 in Eq1 configurations (see Lemma 2.3.3) is $p_W^0 = 1 - u/(2\,\mu_W)$. Because, $u_n \to 2\,\mu_W$, $1 - u_n/(2\,\mu_W) \to 0$.

Using the same arguments, it can also be shown that $p_S^r(n) \to 0$ as $n \to \infty$. Thus, both F_S^n and $F_W^n \to \text{Unif}[0, 2\,\mu_W]$ in distribution as $n \to \infty$.

B.2 Multi-Contestant Equilibria

B.2.1 Marginal return from rank-competition when the number of contestants exceeds 2

In this section, we first show that a lower marginal return from rank competition for strong contestants—the foundation for the greater relative propensity of the strong contestant to

chase the bonus in the manuscript—is a quite general feature of multi-contestant contests when contestants have varying capacities.

Consider a pure risk taking contest, with n contestants, assume that capacity is decreasing in the index of the contestant, i.e., $\mu_1 > \mu_2 > \mu_3 \ldots \mu_{n-1} > \mu_n > 0$. Let $\boldsymbol{\mu} = (\mu_1, \mu_2 \ldots \mu_n)$ represent the vector of capacities.

Assume, as in the manuscript, that contestants submit random bids subject to the capacity constraint—i.e., expected performance is no greater than capacity—and that the contestant with the highest performance receives a rank reward of 1, and all other contestants receive a rank reward of 0.[1] We call such a contest a $\boldsymbol{\mu}$-*contest*. In a given equilibrium of the $\boldsymbol{\mu}$-*contest*, let $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ be the associated vectors of multipliers, and let $\boldsymbol{F}$ represent the vector of contestant performance distributions.

Now consider an all-pay auction in which bidders bid x for an auctioned good. The highest bidder receives the good. All bidders, the winner as well as the losers, pay their bids to the auctioneer. Each of the n bidders has a positive valuation of the good, v_i, $i = 1 \ldots n$. Let $\boldsymbol{v} = (v_1, v_2 \ldots v_n)$ be the vector of such values. We call an all-pay auction where bidders' valuations are given by $\boldsymbol{v}$ a $\boldsymbol{v}$-*auction*. Let $\boldsymbol{G}$ represent the vector of contestant bid distributions in the auction and let $\boldsymbol{u}^* = (u_1^*, u_2^* \ldots u_n^*)$ represent the vector of contestant payoffs given $\boldsymbol{G}$ and $\boldsymbol{v}$.

Lemma B.2.1 *Suppose that $\boldsymbol{F}$ is a $\boldsymbol{\mu}$-contest equilibrium with associated multipliers $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. If*

$$v_i = \frac{1}{\beta_i} \text{ and } u_i^* = \frac{\alpha_i}{\beta_i} \implies \boldsymbol{F} \text{is a } \boldsymbol{v}\text{-auction equilibrium,}$$

and, in this $\boldsymbol{v}$-auction equilibrium, $E[\tilde{x}_i] = \mu_i$.

Proof: The condition for a best reply in the $\boldsymbol{\mu}$-contest is that for all i,

$$\begin{aligned} x \in \text{Supp}(F_i) \Rightarrow \alpha_i + \beta_i x &= \prod_{j \neq i} F_i(x), \\ \text{For all } x \geq 0,\ \alpha_i + \beta_i x &\geq \prod_{j \neq i} F_i(x), \end{aligned} \tag{B.2.76}$$

where $\prod_{j \neq i} F_i(x)$ simply represents the probability of x being the highest performance conditioned on the performance distributions of the other contestants.

[1] Tie bids result in equal division of the rank reward between the tied contestants.

In a $\boldsymbol{v}$-auction, the best reply condition is

$$\begin{aligned} x \in \text{Supp}(G_i) \Rightarrow v_i \prod_{j \neq i} G_i(x) - x = u_i^*, \\ \text{For all } x \geq 0,\ v_i \prod_{j \neq i} G_i(x) - x \leq u_i^*. \end{aligned} \tag{B.2.77}$$

If $\boldsymbol{G} = \boldsymbol{F}$, then, after performing the transformation specified in the lemma, we see that the equilibrium conditions for the $\boldsymbol{v}$-auction are satisfied. Thus F is a $\boldsymbol{v}$-auction equilibrium. Because the distributions of contestant strategies are identical in the contest equilibrium and the corresponding auction equilibrium, it is obvious that the expected contestant performance in contest equilibrium and the bidder's expected bid in the corresponding auction equilibrium are equal to each other. □

Our next lemma shows that, in a $\boldsymbol{\mu}$-contest, the marginal gain from rank competition is always lowest for contestant 1, the contestant with the highest capacity.

Lemma B.2.2 *In any $\boldsymbol{\mu}$-contest equilibrium, $\beta_1 < \min_{j \neq 1} \beta_j$.*

Proof: Suppose that this is not the case, then $\beta_1 \geq \min_{j \neq 1} \beta_j$. In the corresponding auction equilibrium specified in Lemma B.2.1, $v_1 \leq \max_{j \neq 1} v_j$. Let $\mathcal{I} = \{i | v_i = \max v_j\}$.

If $v_1 < \max_{j \neq 1} v_j$, then $1 \notin \mathcal{I}$. We show that this entails a contradiction by considering two cases. First suppose that $\mathcal{I}$ contains only one element, say k, then $v_k > \max_{j \neq k} v_j$. In this case, by Lemma 11 in Baye et al. (1996), F_k strictly stochastically dominates F_j, $j \neq k$. Thus, (a) $\mathbb{E}[\tilde{x}_k] > \mathbb{E}[\tilde{x}_1]$. By Lemma B.2.1, (b) $\mathbb{E}[\tilde{x}_k] = \mu_k$ and $\mathbb{E}[\tilde{x}_1] = \mu_1$. (a) and (b) imply that $\mu_k > \mu_i$, contradicting the definition of the $\boldsymbol{\mu}$ vector, which entails $\mu_1 > \mu_k$.

Next, suppose that $v_1 < \max_{j \neq 1} v_j$ but $\mathcal{I}$ contains more than one element, because, by hypothesis, $1 \notin \mathcal{I}$, Theorem 1 in Baye et al. (1996) shows that the 1's equilibrium auction bid is 0, which is absurd given Lemma B.2.1 and the fact that, by assumption, $\mu_1 > 0$.

Now consider the case where $v_1 = \max_{j \neq 1} v_j$, i.e., $1 \in \mathcal{I}$ but 1 is not the only element in $\mathcal{I}$. In this case, by Theorem 1 in Baye et al. (1996), at least two contestants in $\mathcal{I}$ submit the same bid distribution. This implies that these two bidders' expected bids are the same. But, by assumption, no two components in the $\boldsymbol{\mu}$ vector are identical. And thus, given the correspondence specified in Lemma B.2.1, this is also impossible. □

Lemma B.2.3 *In any $\boldsymbol{\mu}$-contest equilibrium $\beta_2 = \beta_3 = \ldots = \beta_n$, $\alpha_2 = \alpha_3 = \ldots = \alpha_n = 0$, and $\alpha_1 > 0$.*

Proof: Lemma B.2.2 and B.2.1 imply that, in the corresponding auction, $v_1 > \max_{j\neq 1} v_j$. Thus, we can partition the set of contestants as into three sets: $\{1\}$, $\mathcal{W} = \{i|v_i = \max_{j\neq 1} v_j\}$ and $\mathcal{O} = \{i|v_i < \max_{j\neq 1} v_j\}$. Theorem 2 in Baye et al. (1996) shows that if $i \in \mathcal{O}$, then $E[\tilde{x}_i] = 0$. The correspondence established in Lemma B.2.1 shows that $E[\tilde{x}_i] = \mu_i > 0$; thus $\mathcal{O}$ is empty, which implies that $v_2 = v_3 \ldots = v_n$ and thus, by the correspondence, $\beta_2 = \beta_3 = \ldots = \beta_n$. Lemma B.2.2 combined with this result, shows that in the corresponding auction, $v_1 > v_2 = v_3 \ldots = v_n$. Theorem 2 in Baye et al. (1996) shows that in this case $u_1^* > 0$ and $u_j^* = 0$, for $j \neq 1$. Thus because, under the correspondence, $u_i^* = \frac{\alpha_i}{\beta_i}$, $\alpha_1 > 0$ and $\alpha_i = 0$ for $i \neq 1$. □

B.2.2 Example of Atalanta effect in a multi-contestant risk-taking contest

In this section, we present an example of the Atalanta effect in a three-contestant competition. In the competition, submitting the highest performance earns a rank reward of 1. Submitting less than the highest performance earns a rank reward of 0. There are three contestants, S, $W1$, and $W2$ and $\mu_S > \mu_{W1} > \mu_{W2} > 0$. As shown in Lemma B.2.1, the associated multipliers for the three contestants must satisfy the condition that $\alpha_S > 0$, and $\alpha_{W1} = \alpha_{W2} = 0$. Since Lemma B.2.1 also implies that the β's of the two W contestants must be the same, we represent the β's of two W contestants by β_W. Lemma B.2.2 implies that $\beta_S < \beta_W$.

Pure rank competition

The form of equilibrium distributions for the three contestants, (F_S, F_{W1}, F_{W2}) is provided by the following equation,

$$F_{W1} = \begin{cases} \frac{\alpha_S + \beta_S\, x}{\sqrt{\alpha_S + \beta_S\, x_{W2}}} & x \in [0, x_{W2}), \\ \min\left[\sqrt{\alpha_S + \beta_S\, x}, 1\right] & x \geq x_{W2}; \end{cases}$$

$$F_{W2} = \begin{cases} \sqrt{\alpha_S + \beta_S\, x_{W2}} & x \in [0, x_{W2}), \\ \min\left[\sqrt{\alpha_S + \beta_S\, x}, 1\right] & x \geq x_{W2}; \end{cases}$$

$$F_S = \begin{cases} \frac{x\,\beta_W}{\sqrt{\alpha_S + \beta_S\, x_{W2}}} & x \in [0, x_{W2}), \\ \min\left[\frac{x\,\beta_W}{\sqrt{\alpha_S + \beta_S\, x}}, 1\right] & x \geq x_{W2}. \end{cases}$$

Because of the correspondence result, the results in Baye et al. (1996), and Lemma B.2.1, any pure rank competition equilibria must take this form for some parameters, β_S, β_W, x_{W2}, and α_S satisfying

$$0 < \beta_S < \beta_W,\ x_{W2} < \frac{1-\alpha_S}{\beta_S},\ \text{and}\ \alpha_S \in (0,1). \tag{B.2.78}$$

As inspection of Equation (B.2.2) shows that, in any equilibrium, both W contestants place positive mass on 0. The probability that both W contestants place mass on 0 equals α_S. S and $W1$ randomize over the interval $[0,u]$, where $u = \frac{1-\alpha_S}{\beta_S}$. $W2$ randomizes over the interval $[x_{W2},u]$. Over the interval $[x_{W2},u]$, the distribution functions of $W1$ and $W2$ are identical.

Rank and bonus competition

Now consider rank and bonus competition. In addition to the rank reward, the contestants receive a bonus of b if their performance weakly exceeds the bonus threshold r. We aim to verify an equilibrium in which S chases the bonus but $W1$ and $W2$ do not. The form of the candidate equilibrium strategies is given as follows.

$$F_{W1} = \begin{cases} \frac{\alpha_S+\beta_S x}{\sqrt{\alpha_S+\beta_S x_{W2}}} & x \in [0, x_{W2}), \\ \min\left[\sqrt{\alpha_S+\beta_S x}, 1\right] & x \geq x_{W2}; \end{cases}$$

$$F_{W2} = \begin{cases} \sqrt{\alpha_S+\beta_S x_{W2}} & x \in [0, x_{W2}), \\ \min\left[\sqrt{\alpha_S+\beta_S x}, 1\right] & x \geq x_{W2}; \end{cases}$$

$$F_S = \begin{cases} \frac{x\,\beta_W}{\sqrt{\alpha_S+\beta_S x_{W2}}} & x \in [0, x_{W2}), \\ \min\left[\frac{x\,\beta_W}{\sqrt{\alpha_S+\beta_S x}}, 1-p_S^r\right] & x \in [x_{W2}, r), \\ 1 & x \geq r. \end{cases}$$

In the candidate equilibrium, the parameters also satisfy the conditions of Equation (B.2.78) and the parameter, p_S^r, which represents the probability that S targets the bonus, satisfies $p_S^r \in (0,1)$.

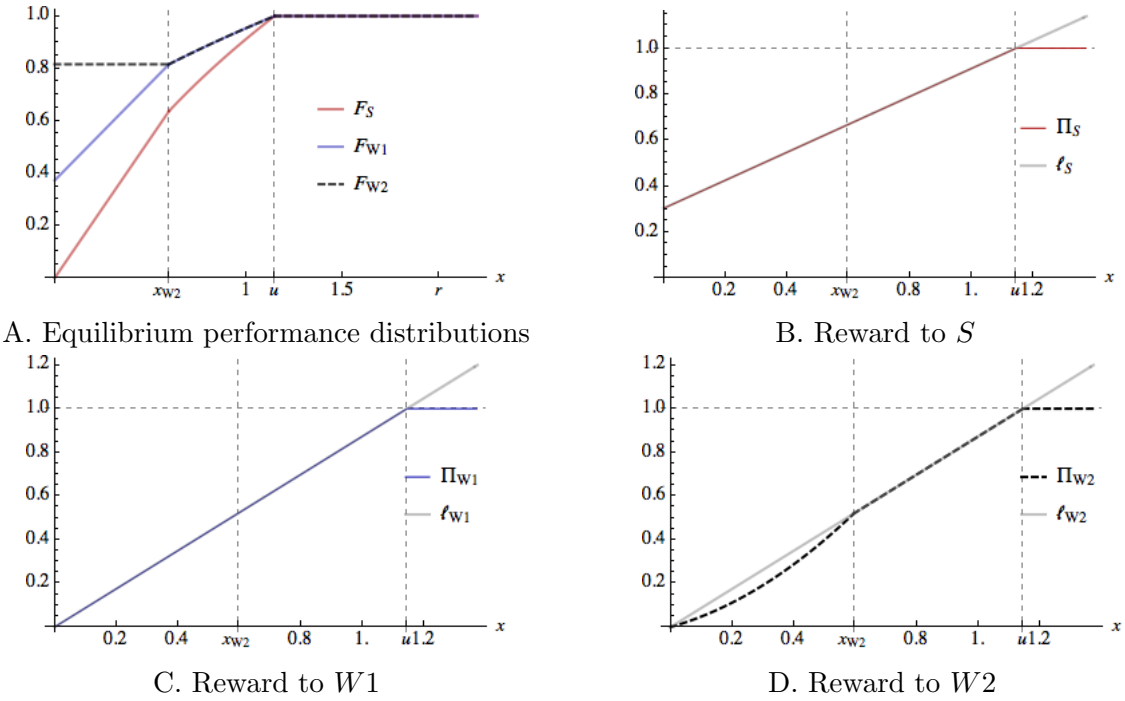
In order to illustrate the effects of introducing bonus compensation in a setting with more than two contestants, we verify equilibria in the pure rank and rank and bonus competition settings and show that qualitatively the behavior of the contestants is quite similar to contestant behavior in the two-contestant setting modeled in the manuscript. In both settings, the parameters—the multipliers and x_{W2}—and, in the case of rank and bonus competition, p_S^r, determine expected performance. However, with more than two

contestants, it is not possible to analytically invert the map between the parameters and expected performance. So we proceed numerically and verify the equilibria for a specific parametric case described in the following table.

Assumed capacity $\mu_S = 0.50$, $\mu_{W1} = 0.2900$, $\mu_{W2} = 0.1584$			
Assumed bonus parameters $r = 2.4$, $b = 1.2477$			
Numerical solution			
Pure rank competition		Rank and bonus competition	
$p^0_{W1} = 0.3726$	$p^0_{W2} = 0.8159$	$p^0_{W1} = 0.2362$	$p^0_{W2} = 0.7766$
$u = 1.1446$	$\alpha_S = 0.3040$	$u = 0.9494$	$\alpha_S = 0.1834$
$\beta_S = 0.6081$	$\beta_W = 0.8737$	$\beta_S = 0.8601$	$\beta_W = 1.0006$
$x_{W2} = 0.5949$	—	$x_{W2} = 0.4880$	$p^r_S = 0.050$

Table B.1: *Parameters for parametric example.*

Plots of the contestant reward functions, Π, and their support lines, ℓ, as well as the candidate equilibrium distributions, for the pure rank competition and rank and the rank and bonus competition are provided by Figures B.2 and B.3 respectively.



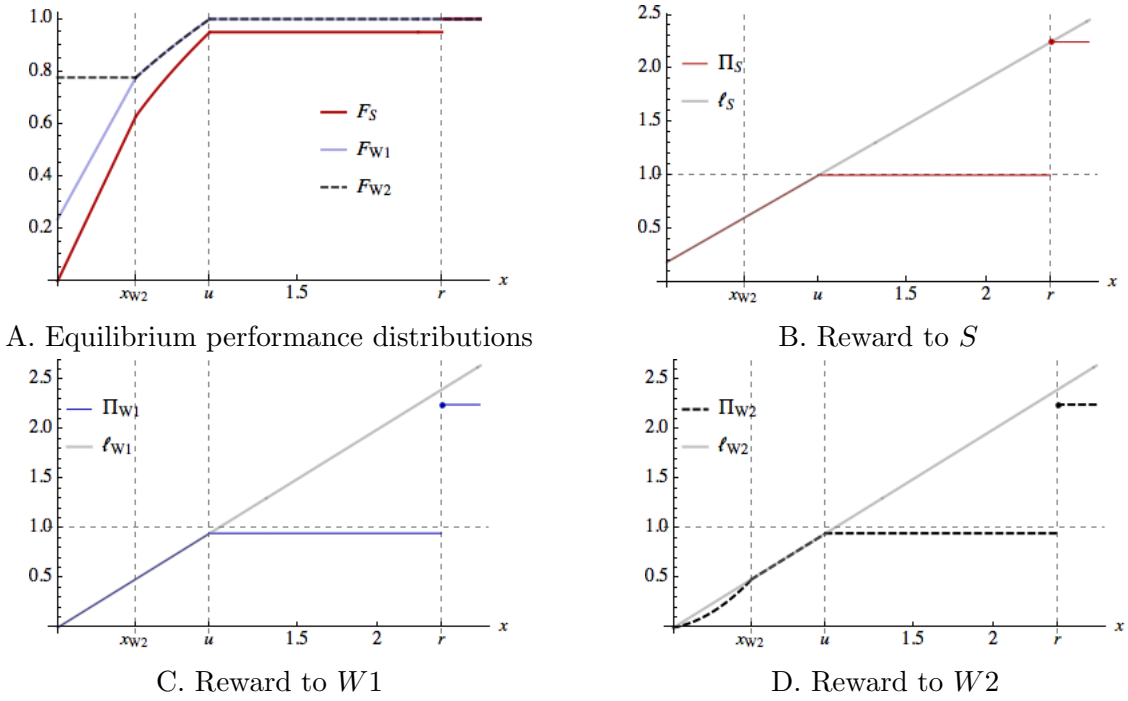
A. Equilibrium performance distributions

B. Reward to S

C. Reward to $W1$

D. Reward to $W2$

Figure B.2: *Pure rank competition: Equilibrium distributions and reward functions.*



A. Equilibrium performance distributions

B. Reward to S

C. Reward to $W1$

D. Reward to $W2$

Figure B.3: *Rank and bonus competition: Equilibrium distributions and reward functions.*

As can be seen from the figures, the best response condition—that all performance levels in the supports of the contestants' distributions lie on their support lines and no performance levels lie above their support lines—are satisfied. Numerical integration shows that the capacity constraint is satisfied with equality for all contestants in the pure rank and rank and bonus competitions. Thus, the candidate equilibria are verified.

Turning to an analysis of the features of the equilibria, we see, from Table B.1, that, qualitatively, equilibrium behavior is quite similar to behavior in the Eq0 configuration developed in the paper: (a) The Atalanta effect is verified, i.e., the probability of downside risk taking for the two W contestants, represented by p^0_{W1} and p^0_{W2}, is reduced by the introduction of bonus compensation. (b) The length of the rank competition region, $(0, u]$, is reduced by the introduction of bonus compensation. Moreover, the logic underlying these two effects is the same as the logic developed in the manuscript: by diverting capacity to bonus chasing, S reduces capacity applied to the rank competition region. This reduction permits $W1$ and $W2$ to compete for rank dominance with S over a smaller region. The smaller region implies that they can satisfy their capacity constraints with performance distributions that place less weight on downside risk taking.

Thus, this example suggests that the qualitative effects of introducing bonus competition into rank competitions between multiple contestants are similar to the qualitative effects with two contestants. The analysis provides no reason to believe that the Atalanta effect depends on the two-contestant setting utilized in the manuscript.

B.3 Dynamic Contests

B.3.1 Dynamic optimization problem

Consider two contestants, F and G, who, as in Section 2.5, compete for rank and bonus rewards in the dynamic contest game. The transition dynamics between the states are provided by the following tree structure, where the first component of the state vector represents the capacity of F and the second represents the capacity of G. By each competition on a time node, the winner will gain a capacity of δ and her rival will lose the same amount, with the sum of capacities constant due to the limited market size. We first specify the contestant value functions and equilibrium distributions at each possible state. This will be the basis for solving the dynamic contest model.

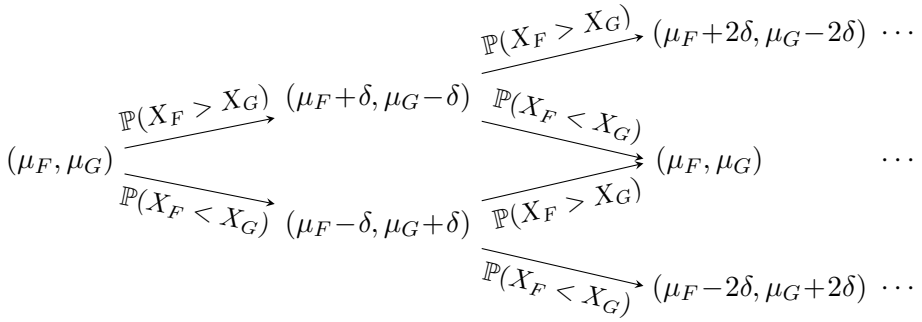


Figure B.4: *State transition dynamics.*

Divide the support of performance into $[0, r)$ where the bonus is not captured, and $[r, \infty)$ where bonus is obtained with certainty. The key insight we use in the formulation of the programming problem, is that picking a performance distribution is equivalent to making the following two decisions: (i) Fix the probability of submitting performance in a specific region and, conditioned on submitting performance in this region, determine how much capacity to allocate to performance in this region. (ii) Conditioned on the allocated capacity and probabilities, determine an equilibrium performance distribution over each region. Using the equilibrium payoffs provided by (ii) we compute the payoffs associated with each allocation and then choose the allocation that maximizes the payoff of a contestant given the other contestant's allocation. We will perform these calculations for contestant F, the payoff calculations for G follow by symmetry.

First consider the payoff to F over $[0, r)$. With probability p_F, F devotes capacity μ_F^L to this region, and her rival, G, targets this region with probability p_G and devotes capacity μ_G^L to this region. Let $\mu_S^L = \max[\mu_F^L, \mu_G^L]$. The leading contestant, i.e., the contestant who possesses more capacity in the current state, will play a uniform distribution over

$[0, u_L] = [0, 2\,\mu_S^L]$; whereas the trailing contestant will randomize between copying his rival and putting a point mass on zero. In order to define the payoff effects of allocations of capacity, the function, N, defined below is quite useful.

$$N(x,y) = \begin{cases} 1 - \frac{1}{2}\frac{y}{x} & \text{if } x > y, \\ \frac{1}{2}\frac{x}{y} & \text{if } x \leq y. \end{cases} \tag{B.3.79}$$

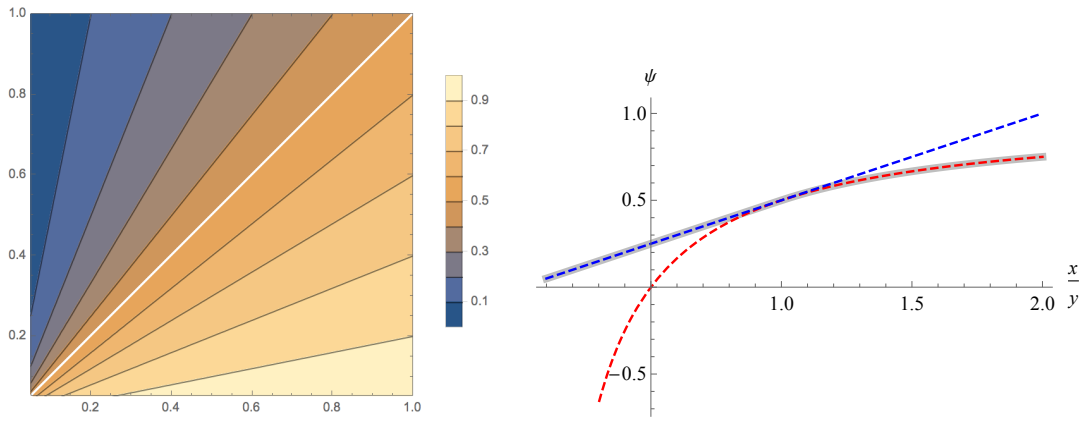
Despite the piecewise definition of N function, we show its nice properties in the following lemma.

Lemma B.3.1 (Properties of N) *Function N is not only continuous but is also differentiable everywhere, both quasiconvex and quasiconcave (i.e., quasilinear), bounded between 0 and 1, and concave in each of its individual arguments.*

Proof: We define a new variable, to be the ratio $r(x,y) \coloneqq x/y$. With the kink point happening at $x = y$ (or equivalently, $r = 1$), the piecewise function (B.3.79) can be rewritten as

$$N(x,y) = \Psi(r) = \begin{cases} 1 - \dfrac{1}{2r} & \text{if } r \geq 1, \\ \dfrac{1}{2}\,r & \text{if } r < 1. \end{cases} \tag{B.3.80}$$

Intuitively, Figure B.5 shows the contouring plot of N with respect to (x,y) and the shape of Ψ with respect to r.



A. Contour plot of $N(\cdot,\cdot)$ B. Function $\Psi(\cdot)$ with respect to r

Figure B.5: *Contouring plot of N with respect to (x,y) and Ψ with respect to x/y.*

By definition, we have

$$\Psi(r) = \frac{1}{2}\int_0^r \min\left[1, \frac{1}{s^2}\right] \mathrm{d}s,$$

where Ψ is differentiable everywhere because $s \mapsto \min[1, 1/s^2]$ is continuous, and $\Psi' > 0$. Thus Ψ is strictly increasing and $\Psi \in [0, 1]$.

The function $r(x, y) = x/y$ is semi-strictly quasilinear. Because Ψ is strictly increasing, $\Psi(x/y) = N(x, y)$ is semi-strictly quasilinear as well.

To put it in detail, let $v_1 = (x_1, y_1)$ and $v_2 = (x_2, y_2)$. Then

$$r\big(\lambda v_1 + (1-\lambda) v_2\big) = \frac{\lambda x_1 + (1-\lambda) x_2}{\lambda y_1 + (1-\lambda) y_2}.$$

If viewed as a function of $\lambda \in [0, 1]$, this function is strictly monotone provided $x_1/y_1 \neq x_2/y_2$. Hence for $\lambda \in (0, 1)$, $r(v_1) \neq r(v_2)$, we have

$$\min\big[r(v_1), r(v_2)\big] < r\big(\lambda v_1 + (1-\lambda) v_2\big) < \max\big[r(v_1), r(v_2)\big].$$

Thus same results apply to the function $N(\cdot, \cdot)$. □

Using N, we can express the probability that contestant F wins in region $[0, r)$ as $N(\mu_F^L, \mu_G^L)$. Winning the rank competition contest only ensures the rank prize if and only if the other contestant, G, also plays in this region. Thus, F will win the rank reward with probability $p_G\, N(\mu_F^L, \mu_G^L)$. Consequently, the expected payoff produced by targeting this region for F will equal the probability that F targets the region, p_F, times the reward above:

$$p_F\, p_G\, N(\mu_F^L, \mu_G^L). \tag{B.3.81}$$

Now consider performance levels in $[r, \infty)$ that aim for both rank and bonus rewards. In this region, the contestant receives the bonus for sure and must always choose performance weakly in excess of r. Thus, we can think of competition in this region involving first paying a participation fee of r and then submitting performance in excess of r. If the conditional capacities allocated to $[r, \infty)$ are μ_F^H and μ_G^H, then essentially the two contestants bid $\mu_F^H - r$ and $\mu_G^H - r$, respectively, for the rank prize. Thus, if we let $\mu_S^H = \max[\mu_F^H, \mu_G^H]$, the support of this region will be $[r, u_H] = [r, r + 2\,(\mu_S^H - r)] = [r, 2\,\mu_S^H - r]$. The payoff for performance targeted at this region consists of a bonus reward and a rank reward. The bonus will always be received. For the rank reward, it is either (i) always attained whenever the rival targets $[0, r)$, or (ii) realized when the rival targets $[r, \infty)$ and is topped. Therefore, again using the N function, the payoff of a contestant, F, by targeting $[r, \infty)$, can be expressed as

$$(1 - p_F)\,\big[\, b + \big(\, p_G + (1 - p_G)\, N(\mu_F^H - r, \mu_G^H - r)\,\big)\,\big]. \tag{B.3.82}$$

Thus, the current reward Π_F to F conditioned on the allocations of capacity by F and G is the following sum

$$\Pi_F \coloneqq p_F\, p_G\, N(\mu_F^L, \mu_G^L) + (1 - p_F)\, \big[\, b + \big(p_G + (1 - p_G) N(\mu_F^H - r, \mu_G^H - r)\big)\, \big].$$

$\mathbb{P}(X_F > X_G)$ denotes the probability that contestant F wins in the game. When F targets $[0, r)$, the winning probability is given by $N(\mu_F^L, \mu_G^L)$. When F targets $[r, \infty)$, the winning probability consists of the sum of two parts: (i) when G targets $[0, r)$ then F wins for sure; and (ii) when both targets $[r, \infty)$ then F wins with probability $N(\mu_F^H - r, \mu_G^H - r)$. Thus, $\mathbb{P}(X_F > X_G) = p_F\, N(\mu_F^L, \mu_G^L)\, p_G + (1 - p_F)\, \big(p_G + (1 - p_G)\, N(\mu_F^H - r, \mu_G^H - r)\big)$. We have $\mathbb{P}(X_F < X_G) = 1 - \mathbb{P}(X_F > X_G)$ by the no-ties property (Remark 2.3.1).

On a given state, each contestant's policy choice can be characterized by a vector $\Theta_i \coloneqq \{p_i, \mu_i^L, \mu_i^H\} \in [0, 1] \times [0, \infty) \times [r, \infty)$, $i = F$ or G. Denote the exogenous value function of F by V_F. The contest can be characterized by two programming problems, with the one for manager F solves

$$\begin{aligned} \max_{\Theta_F} \quad & \Pi_F + \phi\, \big(\, \mathbb{P}\big(\, X_F > X_G\big)\, V_F(\mu_F + \delta, \mu_G - \delta) + \mathbb{P}(X_F < X_G)\, V_F(\mu_F - \delta, \mu_G + \delta)\, \big); \\ \text{s.t.} \quad & p_F\, \mu_F^L + (1 - p_F)\, \mu_F^H \le \mu_F, \\ & \mu_F^H \ge r, \quad p_F \in [0, 1]. \end{aligned}$$

The programming problem for G is completely symmetric. Each contestant solves her programming problem based on the conjectured allocations of the other contestant. Equilibrium results of the conjectured actions of the rival contestant are best responses for the rival.

Reviewing the optimization problem, we find that the capacity constraint is given by the lower contour set of a quasi-convex function. Thus, the capacity constraint set is convex; the other constraint is linear. Hence, the overall constraint set is convex. Although objective function is not quasiconcave, it is differentiable and thus interior maxima for the problem are characterized by the Karush-Kuhn-Tucker (KKT) conditions. We solve the KKT problem analytically for the stationary points and compare these points and payoffs from the boundary solutions. By comparing these local maxima and the boundary solutions we can identify global maximizers for both F and G's optimization problems.

B.3.2 Dynamic equilibrium

We first develop a useful characterization, Lemma B.3.2, of the solution: the ratios of expected "available capacity" allocated to $[0, r)$ and $[r, \infty)$ are the same for both the

leading and trailing contestants.

Lemma B.3.2 (**Equilibrium ratio property**) *In the dynamic contest, a contestant's equilibrium allocation at each state satisfies the following property:*

$$\frac{p_F}{1-p_F}\frac{\mu_F^L}{\mu_F^H-r}=\frac{p_G}{1-p_G}\frac{\mu_G^L}{\mu_G^H-r}, \tag{B.3.83}$$

as long as $p_i \neq 1$, $i = F, G$.

We will prove Lemma B.3.2 after characterizing the equilibrium for exogenous continuation values. The function $N(\cdot,\cdot)$ defined in equation (B.3.79), is a smooth, concave, increasing function of its first argument and a smooth, convex, decreasing function of its second argument, with $\partial_1 N = \frac{1}{2}\min(\frac{1}{y}, \frac{y}{x^2})$ and $\partial_2 N = -\frac{1}{2}\min(\frac{1}{x}, \frac{x}{y^2})$. Because we add in a discounted expected future value, the objective function (NA:PK) changes compared to the static model. The equilibrium for any fixed continuation value can be characterized by the systems provided in the following lemma.

Lemma B.3.3 (**Equilibrium for exogenous continuation values**) *The equilibrium performance at each state of the dynamic game is characterized by the solution to the following equations:*

$$\mu_F = p_F\,\mu_F^L + (1-p_F)\,\mu_F^H, \tag{B.3.84}$$

$$p_G\,\partial_1 N(\mu_F^L,\mu_G^L) = (1-p_G)\,\partial_1 N(\mu_F^H-r,\mu_G^H-r), \tag{B.3.85}$$

$$p_G\,\partial_1 N(\mu_F^L,\mu_G^L) = \frac{(1+A_F)\,p_G\,N(\mu_F^L,\mu_G^L) - \left[b+(1+A_F)\left(p_G+(1-p_G)\,N(\mu_F^H-r,\mu_G^H-r)\right)\right]}{(1+A_F)\,(\mu_F^L-\mu_F^H)}; \tag{B.3.86}$$

$$\mu_G = p_G\,\mu_G^L + (1-p_G)\,\mu_G^H, \tag{B.3.87}$$

$$p_F\,\partial_1 N(\mu_G^L,\mu_F^L) = (1-p_F)\,\partial_1 N(\mu_G^H-r,\mu_F^H-r), \tag{B.3.88}$$

$$p_F\,\partial_1 N(\mu_G^L,\mu_F^L) = \frac{(1+A_G)\,p_F\,N(\mu_G^L,\mu_F^L) - \left[b+(1+A_G)\left(p_F+(1-p_F)\,N(\mu_G^H-r,\mu_F^H-r)\right)\right]}{(1+A_G)\,(\mu_G^L-\mu_G^H)}; \tag{B.3.89}$$

where we denote the terms related to the discounted future payoff as $A_F \coloneqq \phi\left(V_F(\mu_F+\delta,\mu_G-\delta) - V_F(\mu_F-\delta,\mu_G+\delta)\right)$ *and* $A_G \coloneqq \phi\left(V_G(\mu_F-\delta,\mu_G+\delta) - V_G(\mu_F+\delta,\mu_G-\delta)\right)$ *for simplification.*

Proof: Denote the Lagrangian multipliers of contestant F and G by λ_F and λ_G, respectively. The Lagrangian function for contestant F is formulated as follows (Symmetry

implies an analogous expressions for G):

$$\mathcal{L} = p_F\,(1+A_F)\,p_G\,N(\mu_F^L,\mu_G^L)+(1-p_F)\,\big[b+(1+A_F)\,\big(p_G+(1-p_G)\,N(\mu_F^H-r,\mu_G^H-r)\big)\big]- \lambda_F\,\big(p_F\,\mu_F^L+(1-p_F)\,\mu_F^H-\mu_F\big). \quad \text{(B.3.90)}$$

Differentiating the Lagrangian function with respect to F's control variables, we must have the following three equalities at a stationary point:

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial p_F} &= (1+A_F)\,p_G\,N(\mu_F^L,\mu_G^L)-\big[b+(1+A_F)\,\big(p_G+(1-p_G)\,N(\mu_F^H-r,\mu_G^H-r)\big)\big]+\lambda_F\,(\mu_F^H-\mu_F^L)=0,\\
\frac{\partial\mathcal{L}}{\partial \mu_F^L} &= p_F\,\big(\,(1+A_F)\,p_G\,\partial_1 N(\mu_F^L,\mu_G^L)-\lambda_F\,\big)=0,\\
\frac{\partial\mathcal{L}}{\partial \mu_F^H} &= (1-p_F)\,\big(\,(1+A_F)\,(1-p_G)\,\partial_1 N(\mu_F^H-r,\mu_G^H-r)-\lambda_F\,\big)=0.
\end{aligned}$$

Combining with the binding capacity constraint yields the first half of the system, equation (B.3.84)-(B.3.86). Same set of equations characterizes G's performance, equation (B.3.87)-(B.3.89), yielding the final results.

Because of the piecewise definition of function $N(\cdot,\cdot)$, determining the solutions to these stationary points involves analyzing four cases. However, we can focus on just two cases as the other two are obtained through symmetry by exchanging the positions of the two contestants.

(i) $\mu_F^L \le \mu_G^L$, and $\mu_F^H \ge \mu_G^H$:
In this case we have

$$N(\mu_F^L,\mu_G^L)=\frac{\mu_F^L}{2\,\mu_G^L},\qquad N(\mu_F^H-r,\mu_G^H-r)=1-\frac{\mu_G^H-r}{2\,(\mu_F^H-r)}.$$

We first solve the stationary point for $\mu_F^{H\,*}$ and μ_G^{L*}, and the other variables follow:

$$\begin{aligned}
\mu_F^{H\,*} &= \frac{b+(1+A_F)+\lambda_F\,r}{2\,\lambda_F}, & \mu_G^{L*} &= \frac{\lambda_G\,r-b}{2\,\lambda_G},\\
p_F^* &= 1-\frac{2\,\lambda_G\,(\mu_F^{H\,*}-r)}{1+A_G}, & p_G^* &= \frac{2\,\lambda_F\,\mu_G^{L*}}{1+A_F},\\
\mu_F^{L*} &= \frac{2\,\lambda_G\,(\mu_G^{L*})^2}{p_F^*\,(1+A_G)}, & \mu_G^{H\,*} &= \frac{2\,\lambda_F\,(\mu_F^{H\,*}-r)^2}{(1+A_F)\,(1-p_G^*)}+r;
\end{aligned}$$

where λ_F and λ_G take the values that bind the two capacity constraints.

(ii) $\mu_F^L \le \mu_G^L$, and $\mu_F^H \le \mu_G^H$:

In this case we have

$$N(\mu_F^L, \mu_G^L) = \frac{\mu_F^L}{2\,\mu_G^L}, \qquad N(\mu_F^H - r, \mu_G^H - r) = \frac{\mu_F^H - r}{2\,(\mu_G^H - r)}.$$

We can first find the stationary points for the three control variables of G, and then for F, as follows:

$$\begin{aligned}
p_G^* &= \frac{\lambda_F\, r - b}{1 + A_F}, & p_F^* &= 1 - \frac{2\,\lambda_G\,(\mu_G^{H*} - \mu_G^{L*}) - b - \lambda_G\, r}{1 + A_G}, \\
\mu_G^{L*} &= \frac{\lambda_F\, r - b}{2\,\lambda_F}, & \mu_F^{L*} &= \frac{2\,\lambda_G\,(\mu_G^{L*})^2}{(1 + A_G)\,p_F^*}, \\
\mu_G^{H*} &= \frac{1 + A_F + b}{2\,\lambda_F} + \frac{r}{2}, & \mu_F^{H*} &= \frac{2\,\lambda_G\,(\mu_G^{H*} - r)^2}{(1 + A_G)\,(1 - p_F^*)} + r;
\end{aligned}$$

where λ_F and λ_G take the values that bind the two capacity constraints.

The derivation is purely symmetric for the other two cases, (iii) $\mu_F^L \geq \mu_G^L$, and $\mu_F^H \leq \mu_G^H$ and (iv) $\mu_F^L \geq \mu_G^L$, and $\mu_F^H \geq \mu_G^H$. These analytical solutions for the stationary points are indeed the local maximizer because the Hessian matrix is semi-negative definite on the hyperplane orthogonal to the constraint set. By comparing these local maxima and the boundary solutions we can identify the unique global maximizer for both F and G when, for example, $\mu_F \leq \mu_G$, the stationary point for case (i) turns out to produce the highest value. □

Now we give the formal proof of Lemma B.3.2.

Proof: Plugging in the first derivative of $N(\cdot, \cdot)$, equation (B.3.85) gives:

$$p_G \,\min\Big(\frac{1}{\mu_G^L}, \frac{\mu_G^L}{(\mu_F^L)^2}\Big) = (1 - p_G)\,\min\Big(\frac{1}{\mu_G^H - r}, \frac{\mu_G^H - r}{(\mu_F^H - r)^2}\Big).$$

Similarly, from (B.3.88) we have

$$p_F \,\min\Big(\frac{1}{\mu_F^L}, \frac{\mu_F^L}{(\mu_G^L)^2}\Big) = (1 - p_F)\,\min\Big(\frac{1}{\mu_F^H - r}, \frac{\mu_F^H - r}{(\mu_G^H - r)^2}\Big).$$

Hence, we have the following four cases:

(i) $\mu_F^L \leq \mu_G^L$ and $\mu_F^H \leq \mu_G^H$:

$$\frac{p_G}{\mu_G^L} = \frac{1 - p_G}{\mu_G^H - r}, \qquad \frac{p_F\,\mu_F^L}{(\mu_G^L)^2} = \frac{(1 - p_F)\,(\mu_F^H - r)}{(\mu_G^H - r)^2};$$

(ii) $\mu_F^L \leq \mu_G^L$ and $\mu_F^H > \mu_G^H$:

$$\frac{p_G}{\mu_G^L} = \frac{(1-p_G)\,(\mu_G^H - r)}{(\mu_F^H - r)^2}, \qquad \frac{p_F\,\mu_F^L}{(\mu_G^L)^2} = \frac{1-p_F}{\mu_F^H - r};$$

(iii) $\mu_F^L > \mu_G^L$ and $\mu_F^H \leq \mu_G^H$:

$$\frac{p_G\,\mu_G^L}{(\mu_F^L)^2} = \frac{1-p_G}{\mu_G^H - r}, \qquad \frac{p_F}{\mu_F^L} = \frac{(1-p_F)\,(\mu_F^H - r)}{(\mu_G^H - r)^2};$$

(iv) $\mu_F^L > \mu_G^L$ and $\mu_F^H > \mu_G^H$:

$$\frac{p_G\,\mu_G^L}{(\mu_F^L)^2} = \frac{(1-p_G)\,(\mu_G^H - r)}{(\mu_F^H - r)^2}, \qquad \frac{p_F}{\mu_F^L} = \frac{1-p_F}{\mu_G^F - r}.$$

Algebraic simplification shows that these equalities are satisfied. □

Now we numerically characterize the equilibrium of the dynamic predation game. The equilibrium result at the initial state can be found using *dynamic programming*. To formalize these notions in detail, denote each state by $\omega \in \Omega$. Define a transition function, Γ, mapping the current state into the state in the following period, consisting of the following two sub-functions:

$$\Gamma^+(\omega) = \begin{cases} \omega + (\delta, -\delta) & \text{if } X_F > X_G \text{ and } \omega \in \mathbb{R}^2_+, \\ \omega & \text{otherwise;} \end{cases}$$

and

$$\Gamma^-(\omega) = \begin{cases} \omega - (\delta, -\delta) & \text{if } X_F < X_G \text{ and } \omega \in \mathbb{R}^2_+, \\ \omega & \text{otherwise.} \end{cases}$$

Then the Bellman equation for contestant F can be represented by

$$T[V_F](\omega) = \max_{\Theta_F} \Pi_F(\Theta_F, \Theta_G|\omega) + \phi\big[\mathbb{P}(X_F > X_G|\Theta_F, \Theta_G) V_F\big(\Gamma^+(\omega)\big) + \mathbb{P}(X_F < X_G|\Theta_F, \Theta_G) V_F\big(\Gamma^-(\omega)\big)\big], \tag{B.3.91}$$

where Π_F is the function of F's current rewards and ϕ is a discount rate. Operator T is a map from functions to functions. Similarly, for contestant G we have

$$T[V_G](\omega) = \max_{\Theta_G} \Pi_G(\Theta_F, \Theta_G|\omega) + \phi\big[\mathbb{P}(X_G > X_F|\Theta_F, \Theta_G) V_G\big(\Gamma^+(\omega)\big) + \mathbb{P}(X_G < X_F|\Theta_F, \Theta_G) V_G\big(\Gamma^-(\omega)\big)\big]. \tag{B.3.92}$$

Thus, the equilibrium is a fixed point, $V^* = (V_F^*, V_G^*)$, of the Bellman equations (B.3.91) and (B.3.92).

The solution for the fixed point is based on the following iterative procedure: assume a value v_0 at state $(1,1)$ for both F and G. This assumed value will determine the value at all other states, for instance, in the example in the main body of the paper, as follows: once state $(2,0)$ is reached, G is eliminated and thus F receives the rank reward in all subsequent periods. Thus, the value at $(2,0)$ is easy to compute. Symmetry implies that these observations also apply at $(0,2)$ with the roles of the contestants reversed. Values at $(3/2, 1/2)$ and $(1/2, 3/2)$ are given by the discounted expectations of the value at $(1,1)$, and the fixed values at $(0,2)$, and $(2,0)$. Conditioned on these values, we can determine equilibrium performance strategies at each state using Lemma B.3.3. The result will imply a new value v_1 at $(1,1)$, and so forth. Using these iterations, we apply the *contraction mapping* principle to solve for the fixed point solution of the Bellman equation, $V^*(1,1)$, as follows: (i) We start by assigning some value v_0 to state $(1,1)$: We use this value to calculate the values at the other states. (ii) Given these values we solve optimal strategies of the contestants using the conditions in Lemma B.3.3. (iii) These strategies generate a new value at $(1,1)$, $v_1 = T(v_0)$. (iv) We keep iterating steps one and two to get new values $v_{n+1} = T(v_n)$. (v) We continue iteration until we reach the fixed point $v^* = T(v^*) = V^*(1,1)$.

The following figure illustrates the convergence of $\{v_n\}$ to $V^*(1,1)$ in this example for a starting values $v_0 = 0$ or $v_0 = 8$. As implied by the contraction mapping principle, the limit of the sequence is robust to different choices of initial value. For the asymmetric dynamic game, the logic for convergence is similar.

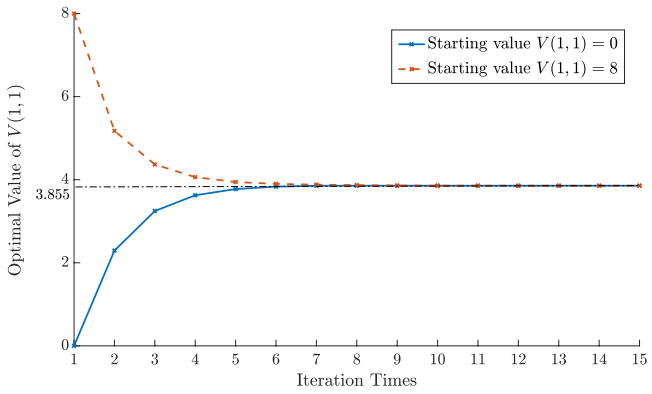


Figure B.6: *Convergent sequences of iterates under different starting values, v_0.*

B.4 Formal Derivations of the Generic Properties of Contest Equilibria

The derivations in this appendix are novel in the sense that they have not been established in a contest game with exactly the same structure as our contest game. However, the arguments and characterizations closely track characterizations for other all pay and risk taking contest games (e.g., Siegel, 2009; Hillman and Riley, 1989). More generally, the approach taken to deriving the results—bounding the support of the payoff distribution with an envelope of affine functions that it majorizes—is equivalent to the concave envelope approach frequently used to analyze all-pay auction, risk-taking, and Bayesian persuasion games (e.g., Aumann et al., 1995; Kamenica and Gentzkow, 2011). We deal with the problem of tied performance, which generates discontinuities in the reward function, by replacing the natural reward function for the game with a more tractable reward function and then establish an equivalence between the equilibria under the natural and tractable functions. This is a very standard approach in games with discontinuous best reply correspondences (cf. Simon and Zame, 1990; Siegel, 2009).

B.4.1 Reward functions, best responses, ties, and support lines

Overall this section is to deal with the problem of tied performance and formally justify the support-line arguments informally developed in the paper. These results will confirm the assertions in Remarks 2.3.2 and 2.3.3 as well as confirming that the reward function used in our definition of equilibrium produces the same set of equilibria as a reward function that splits the rank reward in the event of tied performance.

Payoffs, best replies, and reward functions

Let $-S$ represent W and $-W$ represent S. Consider the following reward functions, Π_i^T and Π_i^N, both defined over $\mathbb{R}_+$:

$$\Pi_i^T(x) = F_{-i}(x-) + \tfrac{1}{2}\big(F_{-i}(x) - F_{-i}(x-)\big) + \mathbb{1}_r(x), \quad i = S, W, \tag{B.4.93}$$

$$\Pi_i^N(x) = F_{-i}(x) + \mathbb{1}_r(x), \quad i = S, W. \tag{B.4.94}$$

Π_i^T accounts for the possibility of tied performance and, in the event of tied performance, splits the rank reward of 1 equally between the two contestants. As an inspection of the following derivations will show, an equal split is not essential to the arguments, any division that does not assign the entire reward to one of the contestants suffices to establish

the subsequent results. An equal division is used here simply to avoid introducing more notation.

In contrast, Π_i^N ignores for the possibility of tied performance and, in the event of tied performance, provides a reward of 1 to both contestants. This is the reward function used in the body of the paper.

Let $\mathcal{P}^+$ be the set of all probability distribution functions supported by $[0, \infty)$ and define $\mathcal{F}_i$ by

$$\mathcal{F}_i = \left\{ F \in \mathcal{P}^+ : \int_{0-}^{\infty} x \, \mathrm{d}F(x) \leq \mu_i \right\}, \quad i = S, W.$$

A best reply for contestant i to F_{-i} under Π^k, $k = T, N$ is a probability distribution F^* that satisfies

$$\int_{0-}^{\infty} \Pi_i^k(x) \, \mathrm{d}F^*(x) = \sup \left\{ \int_{0-}^{\infty} \Pi_i^k(x) \, \mathrm{d}F(x) : F \in \mathcal{F}_i \right\}, \quad k = T, N, \; i = S, W.$$

Both Π_i^T and Π_i^N are bounded, nondecreasing functions defined over $\mathbb{R}_+$ that only differ with respect to how they treat tied performance. Tied performance occurs with positive probability if and only if both contestants choose discontinuous performance distributions that both put positive mass on some performance level. To formalize this notion let

$$\mathcal{M}_i = \{x \geq 0 : F_i(x) - F_i(x–) > 0\}, \quad i = S, W. \tag{B.4.95}$$

Note that because distributions are non decreasing, $\mathcal{M}_i$ is at most countable. We will say that tied performance does not occur if $\mathcal{M}_S \cap \mathcal{M}_W = \varnothing$. When ties do not occur, $\Pi^T = \Pi^N$.

Π^N is right-continuous, and this implies, because it is also non-decreasing, that it is upper semicontinuous. This implies that the map

$$F \hookrightarrow \int_{0-}^{\infty} \Pi^N(x) \, \mathrm{d}F(x)$$

is upper semicontinuous (Ash, 1972, Theorem 4.5.1.b). Thus, whenever the payoff from sequence of distribution converges to the supremum of a contestant's payoff, and the sequence of distributions converges to a distribution function F_o, F_o is a best reply.

In contrast, in general, Π^T is not upper semicontinuous. Thus, the limit of a convergent sequences of distributions producing payoffs that converge to the supremum, need not be a best reply. For this reason, Π^T is not a very convenient reward function. However, it does accurately reflect the fact that the total rank reward is constant, while Π^N does not.

Fortunately, as we will show below, the set of Nash equilibria given Π^T equals the set of Nash equilibria under Π^N which satisfy the condition of no-tied performance. This equivalence results simply because, under Π^T, contestants will never choose in equilibrium to submit performance distributions that result in performance ties with positive probability. This general approach is used in Siegel (2009) to apply Simon and Zame (1990) existence result for games with discontinuous best reply correspondences to all-pay auctions.

Support lines and optimal performance distributions

To establish this equivalence, and to characterize the general properties of equilibrium performance distributions, we require a simple means of characterizing best replies. This desideratum will be supplied by a support line characterization of best replies developed in this section.

Because, in this section, we are only concerned with characterizing the properties of optimal performance distribution for an individual contestant under a given reward function, we will simplify notation by suppressing the subscript representing agent type. Because the results in this section hold for both Π^N and Π^T we will simply represent the contest reward function with Π. The contestant's problem is to maximize her payoff under the contest reward function, Π, over distribution functions in $\mathcal{P}^+$ subject to the capacity constraint, let ν^* represent the supremum of this problem, i.e.,

$$\nu^* = \sup_{F\in\mathcal{P}^+} \left\{ \int_{0-}^{\infty} \Pi(x)\,\mathrm{d}F(x) : \int_{0-}^{\infty} x\,\mathrm{d}F(x) - \mu \leq 0 \right\}. \tag{B.4.96}$$

Because, For all $x \geq 0$, $\Pi(x) \in [0, 1+b]$, $\nu^* < \infty$. Define the associated Lagrange function, $\mathcal{L} : \mathcal{P}^+ \times [0,\infty) \to \mathbb{R}$:

$$\mathcal{L}(F,\lambda) = \int_{0-}^{\infty} \left(\Pi(x) - \lambda\, x\right)\,\mathrm{d}F(x) + \lambda\,\mu. \tag{B.4.97}$$

The objective function,

$$\mathrm{d}F \hookrightarrow \int_{0-}^{\infty} \Pi(x)\,\mathrm{d}F(x)$$

is linear and hence concave, the constraint set is convex, and clearly, there exists F such that the constraint is strictly satisfied. Thus, there exists a multiplier, $\lambda^* \geq 0$, such that

$$\nu^* = \sup_{F\in\mathcal{P}^+} L(F,\lambda^*), \tag{B.4.98}$$

and, if the supremum in (B.4.96) is attained at F^*, then F^* attains the supremum in

equation (B.4.98) (Chapt 1., Theorem 1 Luenberger, 1969). Let

$$\beta = \lambda^*, \quad \text{and } \alpha = \sup_{x \geq 0} \Pi(x) - \beta\, x. \tag{B.4.99}$$

Note that because $\Pi(0) \geq 0$, $\alpha \geq 0$.

Thus, if F^* is an optimal policy, using equations (B.4.97) and (B.4.99), we can express the Lagrange function, evaluated at F^* and λ^*, in terms of α and β as follows:

$$\nu^* = \sup_{F \in \mathcal{P}^+} \mathcal{L}(F, \lambda^*) = \mathcal{L}(F^*, \lambda^*) = \int_{0-}^{\infty} \Big(\Pi(x) - (\alpha + \beta\, x)\Big)\, \mathrm{d}F^*(x) + \alpha + \beta\, \mu. \tag{B.4.100}$$

Inspection of Equation (B.4.100) implies that the optimal performance distribution, F^*, satisfies

$$\begin{aligned} \mathrm{Supp}(F^*) &\subseteq \{x \geq 0 : \Pi(x) - (\alpha + \beta\, x) = 0\}, \\ \forall x \geq 0,\ \Pi^*(x) &\leq \alpha + \beta\, x, \\ 0 &= \int_{0-}^{\infty} \Big(\Pi(x) - (\alpha + \beta\, x)\Big)\, \mathrm{d}F^*(x). \end{aligned} \tag{B.4.101}$$

Clearly $\beta > 0$. If $\beta = 0$ then $\alpha = 1 + b$, which implies, by (B.4.100) and (B.4.101), that $\nu^* = 1 + b$, which is impossible by our assumption that $\mu < r$, and thus the bonus b cannot be captured with probability 1.

Lemma B.4.1 (**Multipliers and support lines**) *If $\Pi = \Pi^T$ or Π^N and F^* is a best response to Π, there exists $\alpha \geq 0$ and $\beta > 0$ and support line, $\ell(x) = \alpha + \beta\, x$, such that F^* satisfies*

$$\mathrm{Supp}(F^*) \subseteq \{x \geq 0 : \Pi(x) = \ell(x)\}, \tag{B.4.102}$$

$$\forall x \geq 0,\ \Pi(x) \leq \ell(x). \tag{B.4.103}$$

Lemma B.4.1 confirms Remark 2.3.3. Note that Lemma B.4.1 establishes necessary conditions for optimal performance distributions under both Π^T and Π^N. These conditions do not speak to the question of whether an optimal performance distribution exists, i.e., whether the supremum of the contestants' optimization problems is attained. In this respect, Π^T and Π^N can differ substantially, as pointed out earlier.

Ties

The next result confirms Remark 2.3.1 by showing that, under Π^T, best responses never produce tied performance. Intuitively, this is obvious, at a tie point, a contestant can

divert infinitesimal to slightly increasing performance at the tie point and to "just top" her rival, breaking the tie and increasing her payoff by a non infinitesimal amount.

Lemma B.4.2 *If F_i^* is best response to Π^T, then $\mathcal{M}_{-i} \cap \text{Supp}_i = \varnothing$, i.e., a best response by i to $-i$ never produces tied performance.*

Proof: Suppose not. Then there exists $x_o \geq 0$ to which both contestants assign positive probability mass. In the event of a tie, the rank reward is divided between the two contestants, with each receiving a rank-based reward of 1/2. Let $\{x_n\}$ be a decreasing sequence converging to x_o.

Consider a contestant's, say S, reward function. The rank based-reward to W if W plays F_W is

$$\Pi_S^T(x_o) = \mathbb{P}[X_W < x_o] + \tfrac{1}{2}\,\mathbb{P}[X_W = x_o], \quad \mathbb{P}[X_W = x_o] > 0. \tag{B.4.104}$$

The rank based reward to S from x_n equals

$$\Pi_S^T(x_n) = \mathbb{P}[X_W < x_n] + \tfrac{1}{2}\,\mathbb{P}[X_W = x_n] \geq \mathbb{P}[X_W \leq x_o] = \mathbb{P}[X_W < x_o] + \mathbb{P}[X_W = x_o]. \tag{B.4.105}$$

The bonus reward at x_n is no less than the bonus reward at x_o. Thus,

$$\Pi_S^T(x_n) \geq \Pi_S^T(x_o) + \tfrac{1}{2}\,\mathbb{P}[X_W = x_o]. \tag{B.4.106}$$

Because $x_o \in \text{Supp}_S$ by hypothesis, condition (B.4.102) implies that $\ell_S(x_o) = \Pi_S(x_o)$. In order for condition (B.4.103) to be satisfied, it must be the case that

$$\forall n \in \mathbb{N}, \quad \ell_S(x_n) \geq \Pi_S^T(x_n). \tag{B.4.107}$$

Equations (B.4.106) and (B.4.107) imply that

$$\forall n \in \mathbb{N}, \quad \ell_S(x_n) \geq \Pi_S^T(x_o) + \tfrac{1}{2}\,\mathbb{P}[X_W = x_o].$$

Thus

$$\Pi_S^T(x_o) = \ell_S(x_o) = \lim_{n\to\infty} \ell_S(x_n) \geq \Pi_S^T(x_o) + \tfrac{1}{2}\,\mathbb{P}[X_W = x_o] > \Pi_S^T(x_o).$$

This contradiction establishes the lemma. □

Lemma B.4.2 provides an equivalence relation between equilibria under Π^T and Π^N which rationalizes the reward function used in the body of the paper (equation (2.3.3)).

Lemma B.4.3 *The following statements are equivalent:*

(i) For $i = S, W$, F_i is a best response to F_{-i} under Π_i^T,
(ii) For $i = S, W$, F_i is a best response to F_{-i} under Π_i^N and $\mathcal{M}_S \cap \mathcal{M}_W = \varnothing$.

Proof: (i) $\Rightarrow$ (ii): If (i) holds, then Lemma B.4.2 implies that $\mathcal{M}_S \cap \mathcal{M}_W = \varnothing$, which implies that $\Pi_i^T = \Pi_i^N$ and thus, (ii) holds.

(ii) $\Rightarrow$ (i): If (ii) holds then $\Pi_i^T = \Pi_i^N$, and thus (i) holds. □

This result formally shows that the payoff function used in the body of the paper, which in the body of the paper is simply called Π and in this appendix, thus far, has been termed Π^N, characterizes equilibrium behavior even when the game specifies a division of the rank reward in the event of ties.

B.4.2 Properties of equilibrium performance strategies

Henceforth, making a slight abuse of notation we represent Π_i^N, $i = S, W$, *simply by* Π_i. Note that this definition coincides with the definition of Π_i in the main body of the paper.

Lemma B.4.4 *If (F_W, F_S) are equilibrium performance distributions,*

(i)
$$(0, r) \cap \text{Supp}_S = (0, r) \cap \text{Supp}_W,$$

(ii)
$$(r, \infty) \cap \text{Supp}_S = (r, \infty) \cap \text{Supp}_W.$$

Proof: We prove (i), the proof of (ii) is omitted because it is virtually identical. Suppose to obtain a contradiction, that there exists x_o such that $x_o \in (0, r) \cap \text{Supp}_W$ but $x_o \notin (0, r) \cap \text{Supp}_S$ (the argument with the roles of the contestants reversed is identical up to transpositions of the type names).

Because, by definition, Supp_S is closed, there exists an open neighborhood N of x_o in $(0, r)$, such that, for all $x \in N$, $x \notin \text{Supp}_S$. Over $(0, r)$, $\Pi_W = F_S$ and thus because F_S is constant on N, Π_W is constant on N.

By hypothesis, $x_o \in N \cap \text{Supp}_W$, thus by condition (B.4.102), $\ell_W(x_o) = \Pi_W(x_o)$. Because ℓ_W is increasing, for $x \in N$ and $x < x_o$, $\ell_W(x) < \ell_W(x_o)$ and, because Π_W is constant over N, $\Pi_W(x) = \Pi_W(x_o)$. Thus, for $x \in N$ and $x < x_o$, $\ell_W(x) < \Pi_W(x)$, contradicting condition (B.4.103). □

Lemma B.4.4 confirms Remark 2.3.2.(a) in the main body of the paper. The next result shows that, in all equilibria, both contestants assign some probability weight to the pure rank competition region.

Lemma B.4.5 *If (F_W, F_S) are equilibrium performance distributions,*

$$(0, r) \cap \text{Supp}_i \neq \varnothing, \quad i = S, W. \tag{B.4.108}$$

Proof: If $(0, r) \cap \text{Supp}_i = \varnothing$, then, by Lemma B.4.4, $(0, r) \cap \text{Supp}_j = \varnothing$, $j \neq i$. Performance distributions that are supported by $[r, \infty)$ are inconsistent with the capacity constraint given the assumption that $\mu_i < r$, $i = S, W$. Thus, both S and W must place positive mass on 0, i.e., the probability of tied performance must be positive. But this is impossible by Lemma B.4.2. □

Lemma B.4.108 will now be used to derive the next lemma, Lemma B.4.6, which confirms Remark 2.3.2.(d).

Lemma B.4.6 *If (F_W, F_S) are equilibrium performance distributions,*

(i)

$$0 \in \text{Supp}_i, \quad i = S, W.$$

(ii)

$$\textit{If } \text{Supp}_i \cap [r, \infty) \neq \varnothing,\ r \in \text{Supp}_i, \quad i = S, W.$$

Proof: Again we prove only (i), because the proof of (ii) is virtually identical. Suppose, to obtain a contradiction, that 0 is not in the support of one of the contestant's performance distributions, say S. Then there would exist a neighborhood of 0, open in $[0, \infty)$ such that for all $x \in N$, $x \notin \text{Supp}_S$. This implies that on N, F_S is constant. Let $x_o = \min\{x \geq 0 : x \in \text{Supp}_S\}$. By Lemma B.4.108, $x_o < r$. For $x \in [0, x_o)$, F_S is constant and thus Π_W is constant. Thus, for all $x \in [0, x_o)$, $\Pi_W(x) = \Pi(0)$, which implies that $\Pi(x_o{-}) = \Pi(0)$. If Π_W is continuous at x_o (i.e., F_S is continuous at x_o) then $\Pi_W(x_o) = \Pi_W(0)$. Condition (B.4.103) implies that $\ell_W(0) \geq \Pi_W(0)$. Because ℓ_W is increasing, and $\Pi(x_o{-}) = \Pi(0)$, $\ell(x_o) > \Pi_W(x_o)$, which, by condition (B.4.102), implies that $x_o \notin \text{Supp}_W$. By Lemma B.4.4 this implies that $x_o \notin \text{Supp}_S$, contradicting the definition of x_o.

Thus, F_S must jump at x_o. This implies by Lemma B.4.2, that F_W does not jump at x_o. On N, F_W is constant, and does not jump at x_o. Hence, $\Pi_S(x_o) = \Pi_S(0)$. Because, by construction, $x_o \in \text{Supp}_S$, $\ell_S(x_o) = \Pi_S(x_o)$ by condition (B.4.102). Because Π_S is constant

on N and ℓ_S is increasing, for $x \in [0, x_o)$, $\ell_S(x) < \Pi_S(x)$, contradicting condition (B.4.103). □

In addition to confirming Remark 2.3.2.(d), Lemma B.4.6 shows that when the rank plus bonus competition region is not empty, its greatest lower bound always equals the bonus threshold. The following lemma shows that the rank and rank plus bonus competition regions are in fact intervals confirming Remark 2.3.2.(b).

Lemma B.4.7 *If (F_W, F_S) are equilibrium performance distributions,*

(i)

$$\text{Supp}_i \cap [0, r) \text{ is connected.}$$

(ii)

$$\text{If } \text{Supp}_i \cap [r, \infty) \neq \varnothing \Rightarrow \text{Supp}_i \cap [r, \infty) \text{ is connected.}$$

Proof: We shall prove (i). The proof of (ii) is virtually identical save for adding the bonus compensation reward b to the contest reward function. To obtain a contradiction suppose that $\text{Supp}_i \cap [0, r)$ is not connected. Without loss of generality, suppose that $i = S$. For $\tau, \nu > 0$, let

$$G = \bigcup_{\tau,\nu>0} \{(x_o - \tau, x_o + \nu) : (x_o - \tau, x_o + \nu) \cap \text{Supp}_S = \varnothing\}. \tag{B.4.109}$$

Let

$$\underline{x} = \inf G, \quad \bar{x} = \sup G.$$

Then $\underline{x}, \bar{x} \in \text{Supp}_S$ and thus, by Lemma B.4.4 and B.4.6, $\underline{x}, \bar{x} \in \text{Supp}_W$. Thus, condition (B.4.102) implies that $\ell_S(\underline{x}) = \Pi_S(\underline{x})$, $\ell_S(\bar{x}) = \Pi_S(\bar{x})$, $\ell_W(\underline{x}) = \Pi_W(\underline{x})$, and $\ell_W(\bar{x}) = \Pi_W(\bar{x})$. Because, for $x < r$, $\Pi_S = F_W$ and $\Pi_W = F_S$ and because G does not meet the supports of S and W, Π_S and Π_W are constant on G. Thus because ℓ_S and ℓ_W are increasing, it must be the case that both F_S and F_W jump up at $\bar{x}$, this implies tied performance with positive probability at $\bar{x}$, contradicting Lemma B.4.2. □

The next two results, Lemma B.4.8 and Lemma B.4.9, are fairly obvious technical results that will be used to establish our final characterization, continuity, in Lemma B.4.10.

Lemma B.4.8 *If (F_W, F_S) are equilibrium performance distributions,*

$$\sup\{\text{Supp}_i \cap [0, r)\} < r.$$

Proof: $\Pi_i(r) > \Pi_i(r-)+b$. Because $b > 0$, condition (B.4.103) implies that $\ell_i(r) \geq \Pi_i(r)$, $\ell_i(r) = \ell_i(r-)$, thus $\ell_i(r) > \Pi_i(r-)$. Because ℓ is continuous, this implies that for all x is a sufficiently small lower neighborhood of r, $\ell_i(r) > \Pi_i(x)$. Thus by condition (B.4.102), such x are not in Supp_i. □

Lemma B.4.9 *If (F_W, F_S) are equilibrium performance distributions,* $\sup(\text{Supp}_i) < \infty$, $i = S, W$.

Proof: We establish this result for S, the proof for W is identical save for transpositions of S and W. For $x \geq r$, $\Pi_S(x) = b + F_S(x) \leq 1 + b$ and $\lim_{x\to\infty} \ell_S(x) = \infty$. So for x sufficiently large, $\ell_S > \Pi_S$, implying, by condition (B.4.103), that for x sufficiently large, $x \notin \text{Supp}_S$. □

Finally, we confirm Remark 2.3.2.(c).

Lemma B.4.10 *If (F_W, F_S) are equilibrium performance distributions, then F_S and F_W are continuous except perhaps at* 0 *and* r.

Proof: Suppose that one of the distributions, say F_S, has a jump at $x_o \neq r$ or 0. Suppose that $x_o < r$. The proof when $x_o > r$ is the same except that the bonus reward is added to the payoffs. In this case, obviously $x_o \in \text{Supp}_S$ which implies, by Lemma B.4.4, that $x_o \in \text{Supp}_W$. For $x < x_o < r$, $\Pi_W = F_S$, thus, by condition (B.4.102), $\ell_W(x_o) = \Pi_W(x_o)$. Lemma B.4.6 shows that $0 \in \text{Supp}_W$, so, again by condition (B.4.102), $\ell_W(0) = \Pi_W(0)$. Because Π_W jumps up at x_o and ℓ is continuous, for a sufficiently small lower neighborhood of x_o, $\Pi_W(x) < \ell_W(x)$. By condition (B.4.102) this implies that points in this neighborhood are not in Supp_W. Thus, there exists $0 < x' < x_o$ such that $x' \notin \text{Supp}_W$ but $0 \in \text{Supp}_W$. $x_o \in \text{Supp}_W$, i.e., Supp_W is not connected, contradicting Lemma B.4.7. □

This result confirms Remark 2.3.2.(c) in the body of the paper. The following lemma simply summarizes the implications of the previous lemmas for equilibrium contestant reward functions.

Lemma B.4.11 *If (F_W, F_S) are equilibrium performance distributions, then there exist constants, α_S, β_S, α_W, β_W, u_H, u_L, such that (a) $\beta_S, \beta_W > 0$ and $\alpha_S, \alpha_W \geq 0$, (b) $u_L \in (0, r)$, (c) $u_H \in [r, \infty)$, and*

(i) If $x \in [0, r)$,

$$\Pi_i(x) = \begin{cases} \alpha_i + \beta_i\, x & x \in [0, u_L], \\ \Pi_i(u_L) & x \in (u_L, r). \end{cases}$$

(ii) If $x \in [r, \infty)$ *and* $u_H = r$,

$$r \in \mathrm{Supp}_S \Leftrightarrow r \notin \mathrm{Supp}_W,$$
$$r \in \mathrm{Supp}_i \Rightarrow \alpha_i + \beta_i\, r = 1 + b, \quad r \notin \mathrm{Supp}_i \Rightarrow \alpha_i + \beta_i\, r \leq 1 + b, \text{ and}$$
$$\Pi_i(x) = 1 + b, \quad i = S, W.$$

(iii) If $x \in [r, \infty)$ *and* $u_H > r$,

$$\Pi_i(x) = \begin{cases} \alpha_i + \beta_i\, x & x \in [r, u_H], \\ 1 + b & x > u_H. \end{cases}$$

B.5 Equilibrium Configurations with Easy Bonus Packages

In this section, we characterize equilibria when bonus packages are easy, i.e., $\mu_S > r$. Our first result is that Eq2 equilibrium configurations can sometimes be sustained when $\mu_S > r$. See the necessary and sufficient condition provided by Lemma B.5.1.

Lemma B.5.1 *When* $\mu_S > r$*, Eq2 equilibria exist if and only if*

$$\mu_S \leq r \left((1 + b) \sqrt{1 + \frac{\mu_W^2}{b\, r\, (b\, r + 2\, \mu_W)}} - b \right). \tag{B.5.110}$$

Proof: We first check that when $\mu_S > r$, no Eq1 configuration exists. This follows easily from showing that

$$1 - p_S^r = 1 - \frac{2\, \mu_S - u}{2\, r - u} = \frac{(r - \mu_S) \left(\mu_W + \sqrt{\mu_W\, (2\, b\, r + \mu_W)} \right)}{r \sqrt{\mu_W\, (2\, b\, r + \mu_W)}} < 0.$$

Thus no Eq1 equilibrium exists.

First, we prove the necessity of condition (B.5.110). For Eq2 equilibrium to be verified, we must have $u_L \geq 0$. By the definition in equation (B.1.24) from Lemma 2.3.4, u_L has

the same sign as

$$b^2\, r\, (r-\mu_S)\, (r+2\,b\,r+\mu_S)+2\,b\, (r-\mu_S)\, (r+2\,b\,r+\mu_S)\, \mu_W+(1+b)^2\, r\, \mu_W^2.$$

Define $x \coloneqq \mu_S - r$. The above expression becomes quadratic in x:

$$-b\,(b\,r+2\,\mu_W)\, x^2-2\,b\,(1+b)\, r\,(b\,r+2\,\mu_W)\, x+(1+b)^2\, r\, \mu_W^2.$$

The positive root of this expression is given by

$$x=\mu_S-r\leq (1+b)\, r\, \left(\sqrt{1+\frac{\mu_W^2}{b\, r\,(b\, r+2\,\mu_W)}}-1\right),$$

i.e., inequality (B.5.110) in the lemma.

Next, we prove the sufficiency, it is sufficient to show that show that if condition (B.5.110) is satisfied, the constraints on other parameters hold.

Start with p_S^r. From equation (B.1.21), for $p_S^r < 1$ we need to show that $\mu_S < r + \mu_W + \mu_W/b$. Condition (B.5.110) is sufficient for this inequality to hold because

$$r+\mu_W+\frac{\mu_W}{b}-r\left((1+b)\sqrt{1+\frac{\mu_W^2}{b\, r\,(b\, r+2\,\mu_W)}}-b\right)=\frac{(1+b)\,(b\, r+\mu_W)\left(1-\frac{b\, r}{\sqrt{b\, r\,(b\, r+2\,\mu_W)}}\right)}{b}>0$$

guarantees that $r + \mu_W + \mu_W/b - \mu_S > 0$.

Next, we show that $u_H > r$. From definition (B.1.23), $u_H - r$ has the same sign as

$$(r-\mu_S)^2(b\, r+\mu_W)^2-\mu_W\,(2\,b\, r+\mu_W)\,(b\, r+\mu_S)^2, \tag{B.5.111}$$

which is a convex function of μ_S defined over $\mu_S \in [r, \overline{\mu}_S]$, where

$$\overline{\mu}_S \coloneqq r\Big((1+b)\sqrt{1+\frac{\mu_W^2}{b\, r\,(b\, r+2\,\mu_W)}}-b\Big).$$

Expression (B.5.111) is negative at $\mu_S = r$. Letting $\mu_S = \overline{\mu}_S$, we see that the expression has the same sign a $1+\frac{b\, r}{b\, r+2\,\mu_W}-2\sqrt{\frac{(b\, r+\mu_W)^2}{b\, r\,(b\, r+2\,\mu_W)}}$, which has the same sign as

$$\left(1+\frac{b\, r}{b\, r+2\,\mu_W}\right)^2-\left(2\sqrt{\frac{(b\, r+\mu_W)^2}{b\, r\,(b\, r+2\,\mu_W)}}\right)^2=-\frac{8\,\mu_W\,(b\, r+\mu_W)^2}{b\, r\,(b\, r+2\,\mu_W)^2}<0.$$

Thus expression (B.5.111) is negative at $\mu_S = \overline{\mu}_S$ as well. Because the maxima of a convex function are attained at its extreme points, $u_H > r$. The constraints for p_S^r, p_W^0, p_S^h and p_W^h follow in like fashion using the definitions provided by (B.1.21) and (B.1.22). □

When an upper bound on μ_S, defined by equation (B.5), is not satisfied, new equilibrium configurations can be verified. Three new configurations are can be realized for some choice of model parameter values.

1. Eq0B: Contestant W randomizes between 0 and uniformly randomizing over the rank plus bonus competition region; contestant S randomizes between r and uniformly randomizing over rank plus bonus competition region.
2. Eq0r: Contestant W randomizes between 0 and r, and contestant S captures both the rank and bonus reward with probability 1. In the Eq0r configuration, S's strategies are not uniformly determined and the capacity constraint for S need not bind; but, if, for any given parameter choice, any performance strategy verifies an Eq0r equilibrium, uniform randomization by S over the rank and bonus competition region verifies an Eq0r equilibrium.
3. EqBB: Contestant W places point mass at r and uniformly randomizes over the rank plus bonus competition region; contestant S uniformly randomizes over rank plus bonus competition region.

Lemma B.5.2 (Eq0B, Eq0r and EqBB)

(i) When $\mu_W \leq r < \mu_S$,

(a) Eq0B equilibria exist if and only if $r\left((1+b)\sqrt{1+\frac{\mu_W^2}{b\,r\,(b\,r+2\,\mu_W)}}-b\right) < \mu_S < r+\frac{r}{2\,b}$, and can be characterized by the following distributions:

$$F_S^* = p_S^r\,\mathbb{1}_r + (1-p_S^r)\,\mathrm{Unif}[r,u], \quad F_W^* = p_W^0\,\mathbb{1}_0 + (1-p_W^0)\,\mathrm{Unif}[r,u].$$

(b) Eq0r equilibria exist if and only if $\mu_S \geq r + \frac{r}{2\,b}$, and can be characterized by the following distributions:

$$F_S^* = \mathrm{Unif}[r, 2\,\mu_S - r], \quad F_W^* = p_W^0\,\mathbb{1}_0 + (1-p_W^0)\,\mathbb{1}_r.$$

(ii) When $r < \mu_W$, EqBB equilibria exist if and only if $\mu_S \geq r + \frac{r}{2\,b}$, and can be characterized by the following performance distributions:

$$F_S^* = \mathrm{Unif}[r, 2\,\mu_S - r], \quad F_W^* = p_W^r\,\mathbb{1}_r + (1-p_W^r)\,\mathrm{Unif}[r, 2\,\mu_S - r].$$

All the parameters are specified in the proof. In all the equilibria, W's capacity constraint binds; In all equilibria, except perhaps Eq0r equilibria, S's capacity constraint binds.

Proof:

(ia) For Eq0B, the capacity constraint for S is $p_S^r\, r + (1 - p_S^r)\,(r/2 + u/2) = \mu_S$. The slope constraint for W, produced by S's distribution, is $\frac{1-p_S^r}{u-r} = \frac{1+b}{u}$. The two constraints jointly determine the value pair

$$p_S^r = \frac{\mu_S - \sqrt{(\mu_S - r)\,(r + 2\,b\,r + \mu_S)}}{r}, \quad u = \frac{b\,r + \mu_S + \sqrt{(\mu_S - r)\,(r + 2\,b\,r + \mu_S)}}{1 + b} > r.$$

To show that $p_S^r < 1$ is equivalent to show that $(\mu_S - r)^2 < (\mu_S - r)\,(r + 2\,b\,r + \mu_S)$, which holds for sure. For $p_S^r > 0$, we need $\mu_S^2 - (\mu_S - r)\,(r + 2\,b\,r + \mu_S) = r\,(r + 2\,b\,r - 2\,b\,\mu_S)$ to be positive. Thus we need the following condition:

$$\mu_S \le r + \frac{r}{2\,b}.$$

The capacity constraint of W is $(1 - p_W^0)\,(r/2 + u/2) = \mu_W$, which implies that

$$p_W^0 = \frac{r + u - 2\,\mu_W}{r + u}. \tag{B.5.112}$$

$p_W^0 \in [0, 1]$ because $u > r > \mu_W$. We also need $p_W^0 \le (1 - b) - (1 - p_W^0)\,\frac{u}{u-r}$ to insure that S's support line lies above $(0, p_W^0)$. Thus we check

$$(1-b) - (1-p_W^0)\,\frac{u}{u-r} - p_W^0 = \frac{b\left(b\,r + \mu_S + \sqrt{(\mu_S - r)\,(r + 2\,b\,r + \mu_S)}\right)^2}{(1+b)^2} - r\,(b\,r + 2\,\mu_W),$$

which is nonnegative if

$$\mu_S \ge r\left((1+b)\sqrt{1 + \frac{\mu_W^2}{b\,r\,(b\,r + 2\,\mu_W)}} - b\right).$$

Now we discuss the Atalanta effect under Eq0B. We have the downside risk taking under pure ranking competition as $p_W^{\text{rank}} = \frac{\mu_S - \mu_W}{\mu_S}$. Thus we compare the downside risk in Eq0B given by (B.5.112) to the one under pure ranking competition,

$$p_W^{\text{rank}} - p_W^0 = \frac{\mu_S - \mu_W}{\mu_S} - \frac{r + u - 2\,\mu_W}{r + u} = -\frac{(r + u - 2\,\mu_S)\,\mu_W}{(r + u)\,\mu_S},$$

which has the same sign as (if we define $x \coloneqq \mu_S - r$)

$$-(r + u - 2\,\mu_S) = \frac{x + 2\,b\,x - \sqrt{x\,(2\,r + 2\,b\,r + x)}}{1 + b},$$

which has the same sign as

$$(x + 2\,b\,x)^2 - x\,(2\,r + 2\,b\,r + x) = -2\,(1 + b)\,x\,(r - 2\,b\,x),$$

which is negative when $\mu_S < r + r/(2b)$. Thus, downside risk is larger under Eq0B configuration compared to the pure ranking game. Thus, the Atlanta effect is reversed.

(ib) For Eq0r, if S randomizes uniformly over a rank plus bonus competition region $[r, u]$, then her capacity constraint implies that $u = 2\,\mu_S - r$. Thus the slope of W's payoff function over $[r, u]$ equals $\frac{1}{2\,(\mu_S - r)}$. The slope of W's support line if W simply chases the bonus equals, $\frac{b}{r}$. For not pursuing rank competition to be a best reply for W we need

$$\frac{b}{r} \geq \frac{1}{2\,(\mu_S - r)}.$$

This holds if $\mu_S \geq r + r/(2\,b)$.

Now we discuss the Atalanta effect under Eq0r. As W plays a two-point discrete strategy, his capacity constraint implies that $p^r_W = \mu_W/r$ and $p^0_W = 1 - \mu_W/r$. By the condition on μ_S, we have

$$p^{\text{rank}}_W = 1 - \frac{\mu_W}{\mu_S} \geq 1 - \frac{\mu_W}{r + \frac{r}{2b}} = \frac{r - \mu_W + \frac{r}{2b}}{r + \frac{r}{2b}} > \frac{r - \mu_W}{r} = p^0_W.$$

Thus, the Atalanta effect persists under Eq0r.

(ii) For EqBB, similarly, the rank plus bonus competition region $[r, u]$ has a support $u = 2\,\mu_S - r$. Contestant W's capacity constraint $p^r_W\,r + (1 - p^r_W)\,\mu_S = \mu_W$ uniquely determines the point mass on bonus threshold, p^r_W. From the assumption $\mu_S > r$, we conclude that $\mu_W > r$.

The support line for S is

$$(1 + b) - (1 - p^r_W)\left(\frac{x - r}{u - r}\right);$$

and for W is

$$(1 + b) - \frac{1}{u - r}\,(u - x).$$

Because the support lines must lie above the origin, condition $\mu_S \geq r + r/(2\,b)$ must hold to sustain the EqBB equilibrium.

As there is no point mass at 0 for EqBB configuration, the Atlanta effect persists in this configuration, albeit for the completely obvious reason that bonus compensation is so large that and so easy to capture that it lifts both contestants minimum performance

to the bonus threshold. The bonus compensation threshold is also so low that bonus compensation, in the absence of rank incentives, does not motivate risk taking. □

We plot the graph of equilibria for easy bonus ($\mu_S > r$) case in Figure B.7. For each bonus package (b, r), one and only one configuration can be sustained.

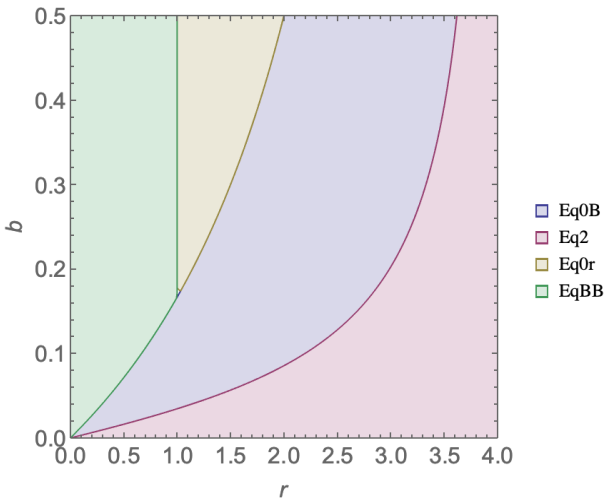


Figure B.7: *Equilibrium configurations and parametric curves when* $\mu_S > r$. In this figure, $\mu_S = 4$ and $\mu_W = 1$.

B.6 Endogenous Bonus Compensation

In this section we assume that contestant capacity is endogenous. More specifically, we assume that each contestant buys a random variable whose cost equals $\gamma_i \mu$, $i = S, W$, and simultaneously selects a distribution satisfying the capacity constraint implied by μ. We assume that the marginal cost of capacity for S, γ_S, is less than the marginal cost of capacity for W, γ_W.

We assume that the welfare function of the designer is linear in contestant capacity and expected bonus payments, i.e.,

$$\alpha \left(\frac{1}{2} \mu_S + \frac{1}{2} \mu_W \right) - \mathbb{E}[\,\text{BonusPmts}\,],$$

where $\alpha > 0$ is a scale parameter. Note that designer welfare is only affected by bonus payments and capacity. Because capacity equals expected performance, the designer is indifferent to the effect of bonus compensation on the riskiness of contestant performance.

The timing of events is as follows: First the designer selects a bonus compensation package, (r, b), where $r \geq 0$ and $b \geq 0$. Note that we do not restrict the designer to challenging

bonus packages, i.e. packages satisfying $r > \mu_S$. Conditioned on the package selected, the two contestants select random performance strategies with mean μ_i, paying capacity cost, $\mu_i\gamma_i$. The structure of the game is identical to the risk-taking game analyzed in the main body of the paper except for the capacity choice.

Note that for a support line $\alpha + \beta\, x$, β measures the gain from relaxing the capacity constraint. Because, in equilibrium, the marginal cost of capacity equals its marginal benefit, and the payoff to each contestant, holding the other contestant's strategies fixed, is a concave function of capacity constraint, optimal capacity choices will satisfy, $\gamma_i = \beta_i$, $i = S, W$. Thus, for each equilibrium configuration, we can determine, using $\gamma = \beta$ relation, the equilibrium mean performance μ_S and μ_W as a function of r, b, γ_S and γ_W. The marginal benefit of capacity, β depends on the equilibrium configuration. The results in Lemma 2.3.5 and in Section B.5, identify the set constraints on the parameters r, b, μ_S, and μ_W required to support each equilibrium configuration. Thus, for each configuration, we can determine an optimal bonus package by solving a constrained maximization problem. Once this is accomplished, we compare welfare under the different configurations. Specifying these constrained optimization problems requires computing the marginal gain from capacity and the expected cost of bonus compensation in the configurations.

B.6.1 Slopes, rewards, and expected bonus payments

In this section, we characterize the slopes of support line (β_i), contestant payoffs evaluating at μ_i ($\Pi_i(\mu_i)$), and the expected bonus payments to both contestants.

The computations for all configurations follow the same pattern. We use the Eq2 configuration to illustrate our approach. In the Eq2 configuration, the slopes for the contestants' support lines are

$$\beta_W = \frac{1+b}{u_H} = \frac{b^2\,(b\,r+\mu_S)^2 + (1+b)^2\,(b\,r+\mu_W)^2}{2\,(b\,r+\mu_S)\,(b\,r+\mu_W)^2},$$

and

$$\beta_S = \frac{1+b-p_W^0}{u_H} = \frac{b^2\,(b\,r+\mu_S)^2 + (1+b)^2\,(b\,r+\mu_W)^2}{2\,(b\,r+\mu_W)\,(b\,r+\mu_S)^2}.$$

From the solutions in Lemma 2.3.4, we can compute the expected bonus payments to contestant S as follows:

$$b\,(p_S^r + p_S^h) = (b\,r+\mu_S)\,b^2\,\frac{b\,r+2\,\mu_W}{2\,(b\,r+\mu_W)^2} + b\,\frac{2\,\mu_S - r\,(1+b^2)}{2\,(b\,r+\mu_S)},$$

and the expected bonus payments to contestant W as follows:

$$b\,p_W^h = b\,\frac{b\,r+\mu_W}{b\,r+\mu_S}\left(\frac{2\,\mu_S\,(1+b)-(1-b^2)\,r}{2\,(b\,r+\mu_S)}-\frac{b^2\,r\,(b\,r+\mu_S)}{2\,(b\,r+\mu_W)^2}\right).$$

Using the same approach, we determine β's and expected bonus payments, using the characterizations of contestant strategies in the main body of the paper for the Eq0, Eq1, and Eq2 configurations and the characterizations of Section B.5 for the other configurations. Configuration Eq0r is not included in this list because the marginal benefit to S from

Equilibria	**Eq0B**	**Eq0r**	**EqBB**	**Eq1**
β_W	$\frac{(1+b)^2}{\mu_S+br+\sqrt{(\mu_S-r)(r+2br+\mu_S)}}$	$\frac{b}{r}$	$\frac{1}{2(\mu_S-r)}$	$\frac{(r-\mu_S)\left(\mu_W+\sqrt{\mu_W(2br+\mu_W)}\right)^2}{2r^2\mu_W\sqrt{\mu_W(2br+\mu_W)}}$
β_S	$\frac{\mu_W\beta_W}{\sqrt{(\mu_S-r)(r+2br+\mu_S)}}$	0	$\frac{\mu_W-r}{2(\mu_S-r)^2}$	$\frac{br+\mu_W+\sqrt{\mu_W(2br+\mu_W)}}{r^2}$
BonusPmt_W	$\frac{2b(1+b)\mu_W}{r+2br+\mu_S+\sqrt{(\mu_S-r)\,(r+2br+\mu_S)}}$	$\frac{b}{r}\,\mu_W$	b	0
BonusPmt_S	b	b	b	$\frac{b\mu_S\left(\mu_W+\sqrt{\mu_W(2br+\mu_W)}\right)-br\mu_W}{r\sqrt{\mu_W\,(2\,b\,r+\mu_W)}}$

Table B.2: *List of support line slopes, β, and expected bonus payments in different equilibrium configurations.*

capacity in this configuration is 0 (S wins both the rank and bonus reward in this equilibrium with probability 1). Thus, it is not possible to equate marginal cost and benefit under this configuration.

B.6.2 Equilibrium expected performance

Using $\gamma=\beta$ relation, the equilibrium mean performance μ_S and μ_W as a function of r, b, γ_S and γ_W, we can now compute expected capacity, μ. Again, taking the Eq2 configuration as our example, the equilibrium mean performance pair (μ_S,μ_W) for the two contestants, S and W, is determined as follows:

$$\begin{cases}\beta_W=\dfrac{b^2\,(b\,r+\mu_S)^2+(1+b)^2\,(b\,r+\mu_W)^2}{2\,(b\,r+\mu_S)\,(b\,r+\mu_W)^2}=\gamma_W,\\ \beta_S=\dfrac{b^2\,(b\,r+\mu_S)^2+(1+b)^2\,(b\,r+\mu_W)^2}{2\,(b\,r+\mu_W)\,(b\,r+\mu_S)^2}=\gamma_S.\end{cases}$$

Thus we have

$$\mu_S=\frac{1}{2}\left(\frac{(1+b)^2}{\gamma_W}+\frac{b^2\,\gamma_W}{\gamma_S^2}-2\,b\,r\right),\qquad \mu_W=\frac{(1+b)^2\,\gamma_S^2+b\,(b-2\,r\,\gamma_S)\,\gamma_W^2}{2\,\gamma_S\,\gamma_W^2}.$$

Similarly, we can determine the equilibrium mean performance for the other configurations. These expected performance levels are provided in Table B.3:

Equilibria	**Eq0**	**EqBB**	**Eq0B**	**Eq1**
μ_W	$\frac{\gamma_S}{2\,\gamma_W^2}$	$r+\frac{\gamma_S}{2\,\gamma_W^2}$	$\frac{\gamma_S\left((1+b)^2-r^2\,\gamma_W^2\right)}{2\,\gamma_W^2}$	$\frac{(b-r\,\gamma_S)^2}{2\,\gamma_S}$
μ_S	$\frac{1}{2\,\gamma_W}$	$r+\frac{1}{2\,\gamma_W}$	$\frac{(1+b)^2-2\,b\,r\,\gamma_W+r^2\,\gamma_W^2}{2\,\gamma_W}$	$r+\frac{\gamma_W\,(b+r\,\gamma_S)}{\gamma_S\,(b-r\,\gamma_S)}$

Table B.3: *List of equilibrium mean performance in different configurations.*

B.6.3 Maximizing designer welfare

We are not able to obtain a closed form solution to the optimal design problem. So we take a numerical approach and provide an example in which Eq1 configuration maximizes the designer welfare. The parameters assumed for this example are $\gamma_W = 0.45$, $\gamma_S = 0.40$, $\alpha = 0.75$. Numerical solutions for each configuration are provided in Table B.4.

Equilibria	Eq0	Eq2	Eq0B	EqBB	**Eq1**
Designer Welfare	0.787	0.787	0.442	0.768	**0.868**

Table B.4: *Designer welfare for different configurations.*

From the table, we see that designer welfare is maximized by a bonus package supporting an Eq1 equilibrium. The optimal bonus package in this configuration is $r = 10.45$, $b = 3.24$. Under the optimal package, the endogenous expected performance levels are $\mu_S = 1.59$ and $\mu_W = 1.11$, and downside risk taking is given by $p_W^0 = 0.045$.

The key to the optimality of Eq1 packages in this example is the relatively low value of α. Numerical analysis of the effect of varying the scale parameter, α, reveals that the fight for optimality between the configurations boils down to a fight between Eq1 and EqBB. The tradeoff characterizing this fight is quite standard: maximizing total output vs. limiting agency rents. EqBB, where bonus compensation ensures that both contestants always submit performance at or above the bonus threshold, requires paying both contestants W's marginal cost of capacity, γ_W. This provides S, who has a lower marginal capacity cost, with significant rents and requires very high expected bonus payments relative to Eq1 optimal solution. However, EqBB solutions allow for unlimited scaling upwards of μ. Thus, if α is sufficiently large, designer welfare is unbounded over EqBB solutions.

The advantage of Eq1 over EqBB configuration is that it pushes the marginal reward to S lower, closer to S's marginal cost of capacity. Thus, the bonus package can be fixed so that

so that S's rents are small. However, the Eq1 solutions do not permit an unlimited scaling up of capacity because Eq1 configurations cannot be supported when bonus rewards are too generous.

This example shows that (a) it is quite easy to construct examples in which challenging bonus packages arise endogenously, as optimal solutions to a designer's welfare maximization problem (b) the welfare function of the designer which engenders challenging bonus package need not depend on the effects of bonus compensation on risk taking.

B.7 Computation of Risk-Averse Agent Welfare

This section explains how the numerical results in Section 2.3.5 were generated. We optimized the utility of a risk-averse agent, who owns the performance of the two contestants, and pays any bonus compensation they receive. The preferences of this agent are defined in Section 2.3.5. Our solution was produced by solving two constrained maximization problems, one for Eq1 configurations and one for Eq2 configurations, over bonus compensation packages. The constraints imposed on these problems were the model parameter assumptions and the necessary and sufficient conditions for the two equilibrium configurations (See the proof of Lemma 2.3.5 in this appendix). For the sake of comparison, we also computed the payoff from pure rank competition and asymptotic limit payoff as downside risk goes to 0.

First, we used Mathematica's "Differential Evolution" algorithm, a global maximization algorithm based on genetic evolution, to identify global maximum points (r^*, b^*) for each configuration, Eq1 and Eq2. Next, using Mathematica's local optimization algorithm, we tested the robustness of our solutions by solving local optimization problems, setting the starting values for these problems equal to the solutions produced by global optimization algorithm. The local maxima identified were equal to the global solutions identified by the global optimization algorithm.

Both optimization exercises produced the same solution, a solution on boundary between the Eq1 and Eq2 regions. The optimal bonus package identified by this optimization exercise yields greater utility for the risk-averse agent than either the limiting downside risk-elimination sequence or pure rank competition. Under the optimal bonus package, downside risk taking is reduced substantially but not eliminated. This exercise shows that the risk profile of contestant behavior produced by introducing bonus compensation can be preferred by a risk averse agent to the risk profile produced by pure rank competition, even though bonus compensation packages are challenging and thus, in isolation, encourage risk taking.

Appendix C

Appendix: Swing Pricing and Fragility in Open-End Mutual Funds

C.1 Tables

Table C.1: *Descriptive statistics on fund characteristics.* This table presents the descriptive statistics for characteristics of corporate bond funds in our sample from January 2006 to December 2016. The unit of observation is fund-month. Panel A shows the descriptive statistics for funds with alternative pricing; Panel B shows the descriptive statistics for funds with traditional pricing. *Flow* is the monthly capital flows into a fund divided by fund's total net assets (in %); *Alpha* is the fund's alpha in the past 12 months (in %); *Size* is natural logarithm of fund's total net assets; *Age* is the natural logarithm of fund age in years; *Expense* is funds' total expense ratio (in %); *Inst* is the fraction of fund's assets held by institutional investors (in %); *Illiquidity* is the value-weighted average of bid-ask spreads of fund's assets. Details on the definitions of the variables are provided in Appendix Chapter A.

Panel A. Alternative Pricing

	Flow	Alpha	Size	Age	Expense	Illiquidity	Inst
P25	-0.6052	-0.0628	17.9023	1.3863	0.5643	0.0054	0.0000
Mean	0.7958	0.2658	18.7737	2.0778	0.8807	0.0094	23.3599
Median	0.0590	0.1948	19.2709	2.1972	0.9218	0.0078	0.0000
P75	1.6364	0.5561	20.1997	2.7081	1.1912	0.0108	42.5579
Std	6.8569	0.5478	2.4715	0.8578	0.4462	0.0072	35.9562

Panel B. Traditional Pricing

	Flow	Alpha	Size	Age	Expense	Illiquidity	Inst
P25	-0.4185	-0.0888	17.7389	1.0986	0.4214	0.0047	0.0000
Mean	1.3315	0.2341	18.7888	1.7591	0.7570	0.0080	34.5601
Median	0.1124	0.1765	18.9854	1.7918	0.7500	0.0072	1.3872
P75	2.1596	0.5450	19.9881	2.3026	1.0200	0.0097	73.7224
Std	7.1247	0.5408	1.7037	0.7749	0.3926	0.0056	40.6099

Table C.2: *Determinants of dilution adjustment factors.* Dependent variable is the daily *Adjustment Factor*, defined as the factor by which the fund NAV is adjusted on a given day. It equals the absolute value of swing factor for swing funds, and equals the half spread in funds' bid and ask prices for dual funds. The unit of observation is fund-day. *Stress* is an indicator variable that equals one if monthly VIX is above the 75^{th} percentile of the sample. *Daily Illiquidity* is the daily value-weighted average of bid-ask spreads of fund's assets; *High Illiquidity* is an indicator variable that equals one for funds with *Daily Illiquidity* above the sample median in a given date. Other fund variables include lagged *Alpha*, *Size*, *Age*, *Expense*, and *Inst*. Appendix Chapter A lists the detailed definitions and calculations of all variables in the regression. Regressions use only swing pricing and dual priced funds. We cluster standard errors by fund and day. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Daily Illiquidity	0.2449***	0.2164***	0.1642***			
	(0.0805)	(0.0798)	(0.0573)			
Stress				0.2411***	0.1404***	0.0010
				(0.0581)	(0.0357)	(0.0193)
High Illiquidity x Stress						0.1930**
						(0.0786)
High Illiquidity						-0.0451
						(0.0295)
Alpha		0.0903*	0.0372		0.0146	0.0248
		(0.0475)	(0.0315)		(0.0254)	(0.0214)
Size		-0.0293*	0.0058		0.0150	-0.0099
		(0.0168)	(0.0187)		(0.0251)	(0.0170)
Age		0.047	0.0204		-0.2311***	-0.1278*
		(0.0509)	(0.1274)		(0.0733)	(0.0751)
Expense		0.0391	0.2541		0.1909	0.3183
		(0.1058)	(0.1843)		(0.1527)	(0.2037)
Inst		-0.0016	0.0034		-0.0057	0.0043*
		(0.0011)	(0.0024)		(0.0037)	(0.0025)
Observations	172,007	133,262	133,262	270,793	199,336	133,262
R-squared	0.077	0.136	0.684	0.022	0.633	0.662
Day FE	Y	Y	Y			
Controls		Y	Y		Y	Y
Fund FE			Y		Y	Y

Table C.3: *Fund flows during market stress.* Dependent variable is *Flow*, defined as the net monthly capital flows into a fund divided by the fund's total net assets. *Alternative* is an indicator variable that equals one if the fund is using one of the alternative pricing mechanisms. *Stress* is an indicator variable that equals one if monthly VIX is above the 75th percentile of the sample. Control variables include lagged values (previous month-end) of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Panel A presents the results of ordinary least squares regressions. In Panel A, the results in columns (1) and (2) are based on the full sample; while those in columns (3) to (7) use the matched sample. Standard errors are clustered by fund and month. Panel B presents the results of quantile regressions for the matched sample. Panel B uses bootstrapped standard errors (estimated through 331 repetitions). * **, *** indicate 10%, 5%, and 1% level of significance, respectively.

Panel A. Least Squared Regressions

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Alternative	-0.7866	-0.7260	-0.6895	-0.0993	-0.6579	-0.3028	-0.8621*
	(0.5297)	(0.5219)	(0.5393)	(0.6608)	(0.5413)	(0.7042)	(0.5157)
Alternative x Stress	1.0410**	0.9934*	1.3711**	1.6676***	1.6369***	1.1131**	1.2982**
	(0.4391)	(0.5589)	(0.5765)	(0.6368)	(0.5876)	(0.5191)	(0.5489)
Stress	-0.9890***	-1.0140***	-1.3467***	-1.7250***		-1.1241***	-1.3075***
	(0.2767)	(0.3688)	(0.3904)	(0.5021)		(0.3832)	(0.3884)
Alpha		0.3526*	0.3212	0.7116***	0.6712**	0.4920**	0.5901***
		(0.1993)	(0.2060)	(0.2190)	(0.3234)	(0.2040)	(0.2094)
Size		0.3001*	0.3164*	-0.6081**	0.3498**	0.0944	0.2648*
		(0.1660)	(0.1679)	(0.2432)	(0.1665)	(0.0902)	(0.1515)
Age		-1.2669***	-1.3062***	-1.0089*	-1.3192***	-1.4305***	-1.0009***
		(0.2811)	(0.2884)	(0.5299)	(0.2820)	(0.2162)	(0.2630)
Expense		0.5694	0.5711	-2.8110***	0.5419	0.3574	-0.3392
		(0.4194)	(0.4294)	(0.9704)	(0.4584)	(0.4003)	(0.4365)
Illiquidity		12.3444	11.3720	52.2817**	-16.1868	22.9051	32.1082
		(24.6152)	(25.9732)	(25.3218)	(28.6746)	(26.2625)	(25.4202)
Inst		-0.0126***	-0.0128***	-0.0278**	-0.0133***	-0.0085*	-0.0143***
		(0.0041)	(0.0040)	(0.0118)	(0.0041)	(0.0046)	(0.0040)
Observations	16,693	10,125	9,670	9,669	9,665	9,670	9,670
R-squared	0.002	0.026	0.026	0.164	0.048	0.057	0.040
Controls	N	Y	Y	Y	Y	Y	Y
Fund FE				Y			
Time FE					Y		
Family FE						Y	
Style FE							Y

Panel B. Quantile Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	q0.5	q1	q5	q10	q25	q50	q75
Alternative x Stress	11.6437***	7.6198***	7.8583***	5.0801***	2.0331***	0.1766**	-0.2902**
	(4.3022)	(2.8641)	(1.3784)	(0.7331)	(0.2843)	(0.0761)	(0.1220)
Stress	-10.3889***	-7.4700***	-6.7904***	-4.6803***	-1.9531***	-0.2071***	0.0460
	(2.6709)	(2.0653)	(1.0768)	(0.6683)	(0.2516)	(0.0660)	(0.0872)
Alternative	-6.2156**	-5.4366***	-4.9201***	-3.2272***	-1.3017***	-0.1058**	0.0293
	(2.5474)	(1.8101)	(0.9338)	(0.5644)	(0.2334)	(0.0517)	(0.0657)
Alpha	4.7913***	1.8043	0.3221	0.8600***	0.5467***	0.1650***	0.1805***
	(1.8070)	(1.3493)	(0.4953)	(0.2251)	(0.1017)	(0.0374)	(0.0446)
Size	-4.9958***	-4.9824***	-1.9402***	-0.8382***	0.0191	0.0847***	0.1536***
	(0.3708)	(0.2724)	(0.1766)	(0.1099)	(0.0450)	(0.0158)	(0.0258)
Age	-3.5406***	-3.6133***	-2.2994***	-1.4509***	-0.6920***	-0.2548***	-0.5550***
	(1.0434)	(0.6806)	(0.4492)	(0.2290)	(0.0862)	(0.0331)	(0.0336)
Expense	-2.8741	0.5738	2.5669***	2.5723***	0.9409***	0.0716	0.0869
	(2.1769)	(1.5059)	(0.7234)	(0.3922)	(0.1677)	(0.0543)	(0.0566)
Illiquidity	-2.0181	29.1972	-44.6820	-21.4796	7.4657	5.2756**	7.9140**
	(95.5381)	(81.9912)	(39.5059)	(16.4298)	(8.0395)	(2.2371)	(3.6732)
Inst	-0.0897***	-0.0748***	-0.0306***	-0.0174***	-0.0097***	-0.0026***	-0.0021***
	(0.0267)	(0.0192)	(0.0075)	(0.0038)	(0.0014)	(0.0004)	(0.0006)
Observations	9,670	9,670	9,670	9,670	9,670	9,670	9,670

Table C.4: *Full swing versus partial swing versus dual priced.* Dependent variable is *Flow*, which is the net monthly capital flows into a fund divided by fund's total net assets; *Stress* is an indicator variable that equals one if monthly VIX is above the 75^{th} percentile of the sample. Using the matched sample, Columns (1) to (3) compare traditionally priced funds to full swing, partial swing, and dual priced funds, respectively, Column (4) uses all. *Full swing* is an indicator variable that equals one if the fund is a full swing fund; *Partial swing* is an indicator variable that equals one if the fund is a partial swing fund; *Dual* is an indicator variable that equals one if the fund is a dual fund. Baseline category in each regression include the matching funds which use the traditional pricing rule. We cluster standard errors by fund and time. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

	(1)	(2)	(3)	(4)
VARIABLES	Partial Swing	Full Swing	Dual	All
Partial swing x Stress	1.3341**			1.3341**
	(0.5487)			(0.5483)
Full swing x Stress		1.3662*		1.3662*
		(0.8062)		(0.8032)
Dual x Stress			2.2829	2.2829
			(1.7599)	(1.7505)
Partial swing	-0.7073			-0.7073
	(0.5804)			(0.5799)
Full swing		-0.8312		-0.8312
		(0.5747)		(0.5726)
Dual			-2.6175**	-2.6175**
			(1.2289)	(1.2223)
Stress	-1.3154***	-1.3154***	-1.3154***	-1.3154***
	(0.3934)	(0.3946)	(0.3952)	(0.3931)
Observations	6,828	3,508	2,682	9,670
R-squared	0.003	0.006	0.014	0.008

Table C.5: *Fund flows during market stress for switchers and their matched funds.* Unit of observation is fund-month. Dependent variable is *Flow*, which is the net monthly capital flows into a fund divided by fund's total net assets. Event period is [-24, 24] months relative to the switching date. *Treated* is an indicator variable that equals one for switching funds; *Stress* is an indicator variable that equals one if monthly VIX is above the 75th percentile of the sample; *Post* is an indicator variable that equals one for the period after the switch. Matching algorithm minimizes the sum of the absolute percentage differences in lagged values of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Matching is performed with replacement. Control variables include lagged values (previous month-end) of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Variable definitions are in Appendix Chapter A. We cluster standard errors by fund and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1)	(2)	(3)	(4)
Stress x Treated x Post	2.4757**	2.6541*	3.5240*	2.6970*
	(1.0303)	(1.4025)	(1.7650)	(1.5681)
Stress x Post	-1.2530	-1.4835	-2.3806	-1.5535
	(0.9583)	(1.2186)	(1.6241)	(1.1058)
Stress x Treated	0.4778	-0.0509	0.2844	0.1874
	(0.6170)	(0.6646)	(0.7919)	(0.6212)
Treated x Post	-2.4801**	-1.5066**	-1.6612*	-1.5362*
	(1.0216)	(0.7094)	(0.8544)	(0.7972)
Post	2.0499**	1.5639**	1.1906	0.8393
	(0.9555)	(0.7033)	(0.7795)	(0.7084)
Treated	-0.3813	-0.5795		-0.6717
	(0.5039)	(0.5636)		(0.6080)
Stress	-0.7108	-0.3546	-0.7276	
	(0.5042)	(0.6225)	(0.7263)	
Alpha		0.3742		0.9244
		(0.3318)		(0.6159)
Size		0.6698***		0.6777**
		(0.2541)		(0.2531)
Age		-1.1006**		-1.1896**
		(0.4946)		(0.5790)
Expense		1.3328*		1.5484*
		(0.7993)		(0.8086)
Illiquidity		28.7498		14.0949
		(32.1364)		(47.5349)
Inst		-0.0153**		-0.0166**
		(0.0064)		(0.0079)
Observations	1,374	1,042	1,374	1,042
R-squared	0.060	0.124	0.276	0.194
Controls	N	Y	N	Y
Fund FE			Y	
Time FE				Y

Table C.6: *End-investor flows during market stress for switchers and their matched funds.* This table shows the effect of alternative pricing rules on end-investor flows during periods of market stress using the sample of switchers and their matched funds. Event period is [-24, 24] months. Matching algorithm is described in the text. Unit of observation is investor-month. Dependent variable is *Flow EndInv*, which is the percentage monthly change in each investor's holding (in number of shares). Columns (1) and (2) show the results for switchers and their matching pairs, respectively; column (3) presents the matched sample results. *Treated* is an indicator variable that equals one for switching funds; *Stress* is an indicator variable that equals one if monthly VIX is above the 75^{th} percentile of the sample; *Post* is an indicator variable that equals one for the period after the switch. Control variables include lagged values (previous month-end) of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Variable definitions are available in Appendix Chapter A. We cluster standard errors by investor and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1) Switchers	(2) Control group	(3) Matched sample
Stress x Treated x Post			0.6341***
			(0.2230)
Stress x Post	0.2596*	-0.3205	-0.3869**
	(0.1346)	(0.2163)	(0.1817)
Stress x Treated			-0.3941**
			(0.1929)
Treated x Post			-0.6698***
			(0.1597)
Post	-0.2127*	0.5194***	0.5219***
	(0.1106)	(0.1376)	(0.1303)
Treated			
Stress	-0.1581**	-0.1020	-0.2525
	(0.0736)	(0.2131)	(0.1789)
Alpha	0.2757***	0.4374**	0.3281***
	(0.1040)	(0.1704)	(0.0856)
Size	-0.5059**	-0.7041***	-0.6739***
	(0.1919)	(0.1294)	(0.1157)
Age	-1.4022*	-1.6378***	-1.7709***
	(0.7129)	(0.4361)	(0.3480)
Expense	-1.8653***	-1.1982	-1.6870***
	(0.6098)	(1.3946)	(0.5755)
Illiquidity	-3.6082	72.1860***	-3.2257
	(8.6290)	(25.3514)	(8.0409)
Inst	-0.0232	-0.0137	-0.0199**
	(0.0139)	(0.0121)	(0.0087)
Observations	251,718	132,675	384,393
R-squared	0.250	0.363	0.338
Investor FE	Y	Y	Y
Controls	Y	Y	Y

Table C.7: *Flow–performance sensitivity.* This table shows the effect of alternative pricing rules on flow–performance sensitivity. Panel A shows the results for the full sample (and their matching pairs) using fund flows. Panel B shows the results for the switching funds (and their matching pairs) using end-investor flows. Matching algorithm is described in the text. Control variables include lagged *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Variable definitions are available in Appendix Chapter A. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

In Panel A, the dependent variable is *Flow*, which is the net monthly capital flows into a fund divided by fund's total net assets. *NegAlpha* equals lagged *Alpha* if it is below zero; it is set to zero otherwise. *Alternative* is an indicator variable that equals one if the fund is using one of the alternative pricing mechanisms. Column (3) presents results for the matched sample.

Panel A. Using Fund Flows for the Full Sample

	(1)	(2)	(3) Matched Sample
VARIABLES	Fund Flow	Fund Flow	Fund Flow
NegAlpha		5.8227***	7.0479***
		(1.4523)	(1.8523)
NegAlpha x Alternative		-4.0730***	-5.0280***
		(1.4817)	(1.8578)
Alpha	1.5287***	0.2767	0.1114
	(0.5412)	(0.5690)	(0.6177)
Alpha x Alternative	-0.5253	0.2639	0.4354
	(0.4838)	(0.5415)	(0.5797)
Alternative	-0.3690	-0.8427	-0.9280*
	(0.5165)	(0.5441)	(0.5530)
Size	0.2743*	0.2766*	0.3005**
	(0.1459)	(0.1465)	(0.1494)
Age	-1.0158***	-1.0127***	-1.0070***
	(0.2576)	(0.2552)	(0.2630)
Expense	-0.3771	-0.2997	-0.3126
	(0.4647)	(0.4572)	(0.4695)
Illiquidity	12.3976	22.8068	20.3520
	(27.0505)	(27.5163)	(28.2355)
Inst	-0.0149***	-0.0138***	-0.0142***
	(0.0040)	(0.0039)	(0.0039)
Observations	10,125	10,125	9,670
R-squared	0.060	0.063	0.064
Time FE	Yes	Yes	Yes

In Panel B, the dependent variable is *Flow EndInv*, which is percentage monthly change in each investor's holding (in number of shares). *Treated* is an indicator variable that equals one for switching funds; *Post* is an indicator variable that equals one for the period after the switch. *Alpha* is the fund's alpha in the past 12 months. Event period is [-24, 24] months. *NegAlpha* equals lagged *Alpha* if the fund's lagged *Alpha* is negative (or below the 25th percentile, in column 3); it is set to zero, otherwise. Regressions include the interaction terms of *Alpha* (and *NegAlpha*) with *Treated* and *Post*. We cluster standard errors by investor and month.

Panel B. Using End-Investor Flows for Switchers and their Matched Funds

VARIABLES	(1) Flow EndInv	(2) Flow EndInv	(3) Flow EndInv
NegAlpha x Treated x Post		-1.5247*	-4.5641**
		(0.8013)	(1.8491)
NegAlpha x Post		1.3472*	4.4680**
		(0.7123)	(1.8426)
NegAlpha x Treated		-0.3741	0.6165
		(1.6914)	(1.8189)
NegAlpha		0.4240	0.5432*
		(0.6908)	(0.3189)
Alpha x Treated x Post	0.0180	0.2071	0.1053
	(0.1364)	(0.1588)	(0.1477)
Alpha x Post	0.0178	-0.1013	-0.0392
	(0.1268)	(0.1297)	(0.1220)
Alpha x Treated	-0.1445	-0.1122	-0.1750
	(0.1138)	(0.1208)	(0.1185)
Alpha	0.4578***	0.3965***	0.4675***
	(0.1059)	(0.0989)	(0.0977)
Treated x Post	-0.4371***	-0.5233***	-0.4847***
	(0.1010)	(0.1146)	(0.1075)
Post	0.4163***	0.4475***	0.4411***
	(0.1000)	(0.1000)	(0.0980)
Size	-0.6631***	-0.6537***	-0.6675***
	(0.0683)	(0.0697)	(0.0696)
Age	-1.7081***	-1.6450***	-1.6804***
	(0.2161)	(0.2202)	(0.2205)
Expense	-1.4161***	-1.4042***	-1.3883***
	(0.1848)	(0.1868)	(0.1852)
Inst	-0.0187***	-0.0181***	-0.0186***
	(0.0057)	(0.0057)	(0.0057)
Illiquidity	-1.7149	-0.9046	-1.7574
	(1.6117)	(1.8760)	(1.8032)
Observations	384,393	384,393	384,393
R-squared	0.338	0.338	0.338
Investor FE	Y	Y	Y
Controls	Y	Y	Y

Table C.8: *Fund exit.* Dependent variables are *Exit* defined as an indicator variable that is equal to one if the fund exits the market during our sample period, and equal to zero if the fund remains alive. *Merged* is an indicator variable that equals one if the fund has become obsolete due to a merger event during the sample period and equals zero if the fund remains alive. *Liquidated* is an indicator variable that equals one if the fund has become obsolete due to a liquidation during our sample period, and zero if the fund remains alive. *Alternative* is an indicator variable that equals one if the fund is using one of the alternative pricing mechanisms. The regression model in column (7) uses the matched sample including the control variables of lagged *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Appendix Chapter A lists the detailed definitions and calculations of all variables in the regression. The unit of observation is fund-month. We cluster standard errors by fund and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1) Exit	(2) Merged	(3) Liquidated	(4) Exit	(5) Merged	(6) Liquidated	(7) Liquidated Matched sample
Alternative	-0.0474	0.0611*	-0.1081*	-0.1229	0.0134	-0.1723**	-0.0885*
	(0.0599)	(0.0312)	(0.0560)	(0.0884)	(0.0472)	(0.0836)	(0.0479)
Alpha				0.0527	0.0425	0.0354	0.0423
				(0.0567)	(0.0508)	(0.0441)	(0.0453)
Size				-0.0422***	-0.0283***	-0.0256**	-0.0214**
				(0.0119)	(0.0105)	(0.0104)	(0.0095)
Age				0.0057	0.0395*	-0.0345	-0.0269
				(0.0256)	(0.0211)	(0.0211)	(0.0198)
Expense				0.1853***	0.0788	0.1703***	0.1492**
				(0.0649)	(0.0543)	(0.0590)	(0.0592)
Illiquidity				25.7466***	27.3780***	4.5239	4.2281
				(5.1310)	(4.3891)	(7.6024)	(7.7906)
Inst				0.0017***	0.0010*	0.0015***	0.0017***
				(0.0006)	(0.0006)	(0.0006)	(0.0006)
Constant	0.1739***	0.0256	0.1556***	1.0122***	0.5262**	0.8208***	0.6486***
	(0.0561)	(0.0254)	(0.0542)	(0.2819)	(0.2425)	(0.2753)	(0.2243)
Observations	299	281	277	174	167	163	159
R-squared	0.003	0.006	0.026	0.569	0.535	0.414	0.416
Controls	N	N	N	Y	Y	Y	Y
Family FE	N	N	N	Y	Y	Y	Y

Table C.9: *Cross-fund differences.* Dependent variable is *Flow*, defined as the net monthly capital flows into a fund divided by the fund's total net assets. *Alternative* is an indicator variable that equals one if the fund is using one of the alternative pricing mechanisms. *Stress* is an indicator variable that equals one if monthly VIX is above the 75^{th} percentile of the sample. Regressions use the matched sample including the control variables of lagged *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst* (*Retail* in column (3)). Column (1) introduces interaction terms with lagged *Illiquidity*; column (2) with lagged *Ownership Concentration*, which is the Herfindahl–Hirschman Index of end-investors' ownership; column (3) with *Retail*, which is $1 - Inst$. Appendix Chapter A lists the detailed definitions and calculations of all variables in the regression. The unit of observation is fund-month. We cluster standard errors by fund and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1)	(2)	(3)
Alternative x Stress x Illiquidity	26.7550*		
	(14.5103)		
Stress x Illiquidity	-28.7799*		
	(17.2803)		
Alternative x Stress x Ownership Concentration		-6.0387*	
		(3.5858)	
Stress x Ownership Concentration		3.9060	
		(3.0725)	
Alternative x Stress x Retail			0.0243**
			(0.0114)
Stress x Retail			-0.0121
			(0.0098)
Alternative x Stress	1.9128**	2.4103***	2.2904**
	(0.9164)	(0.6379)	(0.9016)
Stress	-1.8045***	-1.8842***	-1.9369**
	(0.6519)	(0.6172)	(0.7769)
Alternative x Illiquidity	-88.0525		
	(61.1618)		
Alternative x Ownership Concentration		6.1098***	
		(2.2760)	
Alternative x Retail			-0.0101
			(0.0118)
Alternative	-0.8307*	-1.8383***	-1.1000
	(0.4745)	(0.3880)	(0.9592)
Illiquidity	78.7570		
	(59.6897)		
Ownership Concentration		-6.8252***	
		(2.0888)	
Retail			0.0194*
			(0.0101)
Observations	9,670	8,303	9,670
Controls	Y	Y	Y
R-squared	0.031	0.027	0.026

Table C.10: *End-investor flows during market stress: Role of investor characteristics.* The dependent variable, *Flow EndInv*, is percentage monthly change in each investor's holdings (number of shares). *Treated* is an indicator variable that equals one for switching funds; *Post* is an indicator variable that equals one for the period after the switch. *Inst Investor* is an indicator variable that equals one if the end-investor is an institutional client (set to zero otherwise). *Patient Investor* is an indicator variable that equals one if the end-investor has an investment horizon above the sample median (set to zero otherwise). Investment horizon is the number of months the investor holds his shares after an initial purchase. In calculating investment horizon, we use purchases before December 2014. Columns (1), (3), (4), (5), and (7) present the results for switching funds; columns (2) and (6) present the results for control funds. We cluster standard errors by investor and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1) Treated	(2) Control	(3) Treated Retail-oriented	(4) Treated Inst-oriented	(5) Treated	(6) Control	(7) Treated
Stress x Post x Inst Investor	0.4409**	0.1051	0.5693***	0.0074*			0.4604**
	(0.1970)	(1.1571)	(0.1841)	(0.0039)			(0.1992)
Stress x Post	0.1275	0.3316	0.3425	0.0009	0.1711	-0.2283	0.0564
	(0.1336)	(0.9681)	(0.2570)	(0.0689)	(0.1185)	(0.2295)	(0.1170)
Post x Inst Investor	-	-	-	-			-
Stress x Inst Investor	-0.2614*	-1.9006*	-0.3700***	-0.1270			-0.2604*
	(0.1351)	(1.0350)	(0.1236)	(0.2008)			(0.1360)
Stress x Post x Patient Investor					0.1249**	-0.1198	0.0881*
					(0.0581)	(0.2970)	(0.0446)
Post x Patient Investor					-	-	
Stress x Patient Investor					-0.0899	0.1691	0.0091
					(0.0563)	(0.2887)	(0.0363)
Post	-0.2025*	0.2645	-0.4499**	0.0074	-0.2160*	0.5246***	-0.2017*
	(0.1101)	(0.3958)	(0.2079)	(0.0584)	(0.1112)	(0.1382)	(0.1091)
Stress	-0.0541	-0.5807	-0.2204	-0.0178	-0.0975*	-0.0172	-0.0599
	(0.0620)	(0.7479)	(0.1986)	(0.0132)	(0.0543)	(0.2004)	(0.0468)
Inst Investor	-	-	-	-	-	-	-
Patient Investor	-	-	-	-	-	-	-
Observations	231,305	64,513	145,604	85,701	251,718	132,675	231,305
R-squared	0.250	0.404	0.284	0.187	0.251	0.363	0.250
Investor FE	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y
	p-value=0.071		p-value=0.066		p-value=0.064		

Table C.11: *Fund flows and future fund performance.* Dependent variable is the abnormal fund return in month $t+1$, calculated as the difference between fund's return (calculated using unadjusted fund prices) and fund's exposure to global bond market and global stock market returns. Fund's exposure to global bond market and global stock market returns are calculated as $\beta 1_{t\to t-11} \times Bond\ market\ return_{t+1}$ and $\beta 2_{t\to t-11} \times Stock\ market\ return_{t+1}$. *Net Outflow* is the net monthly outflows in t, which equals *Flow* if $Flow < 0$, and it equals zero if $Flow \geq 0$. *Net Inflow* is the net monthly inflows in t, which equals *Flow* if $Flow > 0$, and it equals to zero if $Flow \leq 0$. *Alternative* is an indicator variable that equals one if the fund is using one of the alternative pricing mechanisms. Control variables include year-month fixed effects, as well as *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst* measured as of time t. Appendix Chapter A lists definitions of all variables in the regression. Columns (1) and (3) report results for the full sample; columns (2) and (4) report results for the subsample of funds with more illiquid assets (*Illiquidity* above sample median). The unit of observation is fund-month. We cluster standard errors by fund and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1) Full Sample	(2) High Illiquidity	(3) Full Sample	(4) High Illiquidity
Net Outflow	-0.0352**	-0.0546*		
	(0.0170)	(0.0300)		
Net Outflow x Alternative	0.0372**	0.0662**		
	(0.0184)	(0.0317)		
Net Inflow			0.0028	0.0079
			(0.0081)	(0.0111)
Net Inflow x Alternative			-0.0019	-0.0101
			(0.0117)	(0.0160)
Alternative	-0.0313	-0.0161	-0.0019	0.0461
	(0.0557)	(0.0589)	(0.0580)	(0.0679)
Size	-0.0157	0.0087	-0.0161	0.0082
	(0.0098)	(0.0128)	(0.0100)	(0.0128)
Age	0.0400	0.0099	0.0382	0.0109
	(0.0442)	(0.0467)	(0.0437)	(0.0454)
Expense	-0.2079***	-0.2039***	-0.2104***	-0.2122***
	(0.0792)	(0.0784)	(0.0783)	(0.0766)
Illiquidity	1.7642	-0.1276	1.7650	-0.2475
	(6.2322)	(7.3500)	(6.2592)	(7.3619)
Inst	0.0006	0.0006	0.0006	0.0006
	(0.0007)	(0.0008)	(0.0007)	(0.0008)
Observations	7,827	4,146	7,827	4,146
R-squared	0.415	0.480	0.415	0.479
Month-Year FE	Y	Y	Y	Y

Table C.12: *Tracking error and new investors.* In column (1) and (2), the dependent variable is *Tracking Error* defined as -1 times the R-squared obtained from the rolling 12-month one-factor regression; in column (3), the dependent variable is *Flow* defined as the net monthly capital flows into a fund divided by the fund's total net assets.; in column (4), the dependent variable is *New Investors* defined as the number of a fund's new investors divided by the fund's total number of investors in each month. Columns (1) to (3) use the full sample; column (4) uses periods outside stress. Control variables include lagged (previous month-end) values of *Alpha*, *Size*, *Age*, *Expense*, *Inst*, and *Illiquidity*. Definitions of variables are in Appendix Chapter A. We cluster standard errors by fund and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

	(1)	(2)	(3)	(4)
VARIABLES	Tracking Error	Fund Flow	Fund Flow	New Investors
Alternative	0.0989***	-0.3721	-0.0318	-0.8640**
	(0.0306)	(0.5098)	(0.5122)	(0.4247)
Tracking Error			-3.2157**	
			(1.4418)	
Alpha	0.0153	0.6748**	0.7278**	0.4183**
	(0.0109)	(0.3133)	(0.3213)	(0.2084)
Size	-0.0038*	0.3214**	0.3085*	0.1150
	(0.0020)	(0.1638)	(0.1624)	(0.0829)
Age	-0.0037	-1.2900***	-1.3029***	-1.5679***
	(0.0085)	(0.2791)	(0.2777)	(0.2288)
Expense	-0.0448*	0.5208	0.4037	0.5554
	(0.0231)	(0.4461)	(0.4607)	(0.3911)
Inst	-0.0007***	-0.0135***	-0.0155***	-0.0036
	(0.0002)	(0.0041)	(0.0043)	(0.0045)
Illiquidity	4.0906***	-12.0907	0.9509	40.1989
	(1.0807)	(27.9568)	(28.6336)	(37.2428)
Observations	10,604	10,125	10,125	7,259
R-squared	0.257	0.045	0.047	0.087
Controls	Y	Y	Y	Y
Month-Year FE	Y	Y	Y	Y

Table C.13: *Pricing rules and fund portfolio adjustments.* This table shows the effect of alternative pricing rules on fund's cash holdings (column 1), asset concentration (column 2) and load charges (column 3 and 4). *Cash* is fund's total cash holdings (including cash equivalents) divided by fund's total assets; *Asset Conc* is Herfindahl–Hirschman Index of fund's asset holdings in each month; *Front Load* is the value-weighted average of (minimum) front load charges across share classes of a given fund; *Back Load* is the value-weighted average of (minimum) rear load charges across share classes of a given fund. Variable definitions are in Appendix Chapter A. *Alternative* is an indicator variable that equals one if the fund is using one of the alternative pricing mechanisms. Control variables include lagged values (previous month-end) of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. We cluster standard errors by fund and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

	(1)	(2)	(3)	(4)
VARIABLES	Cash	Asset Conc	Front Load	Back Load
Alternative	-3.2586**	-0.0051	-0.0679	-0.0972
	(1.2938)	(0.0123)	(0.3429)	(0.0837)
Alpha	-0.1688	-0.0181*	-0.1563	0.0092
	(0.5638)	(0.0094)	(0.1595)	(0.0100)
Size	-0.0470	-0.0010	-0.0335	-0.0070
	(0.2794)	(0.0019)	(0.0772)	(0.0060)
Age	-0.7100	0.0059	-0.0279	0.0135
	(0.5017)	(0.0068)	(0.1641)	(0.0189)
Expense	0.9906	0.0159	2.4685***	0.0159
	(1.3146)	(0.0100)	(0.3059)	(0.0375)
Inst	-0.0090	0.0000	0.0125***	-0.0005
	(0.0115)	(0.0001)	(0.0039)	(0.0005)
Illiquidity	13.5219	-1.1383**	20.1502	-1.5599
	(64.2152)	(0.4474)	(19.3040)	(1.1896)
Observations	9,158	10,563	10,254	10,254
R-squared	0.278	0.039	0.204	0.050
Controls	Y	Y	Y	Y
Time FE	Y	Y	Y	Y

C.2 Online Tables

Table C.14: *Swing thresholds of partial swing funds.* This table shows the frequency distribution table for swing thresholds used by partial swing funds in our sample. *Threshold* is the swing threshold (in absolute terms) used by partial swing funds. *Frequency* is defined in %. For funds with multiple thresholds (around 1% of partial swing funds), we report the minimum.

Threshold	Frequency
0.01%	4.59
0.50%	3.1
1%	40.36
1.50%	1.17
2%	4.05
2.50%	2.29
3%	34.39
4%	1.2
5%	6.92
6%	0.22

Table C.15: *Extended robustness tests.* Dependent variable is *Flow*, defined as the net monthly capital flows into a fund divided by the fund's total net assets. *Alternative* equals one if the fund is using one of the alternative pricing mechanisms. In columns (1) to (5), *Stress* equals one if monthly VIX is above the 75th percentile of the sample. Column (1)-(4) introduces fixed effects of *Region of Sale*, *Domicile*, *Investment Objective*, *Investment Area*. Column (5) includes front-end and rear-end load charges. Columns (6) to (8), we use alternative definitions of market stress. *Stress* is defined according to the 75th percentile of TED spread, LIBOR, and Merrill Lynch's MOVE index, respectively. In column (9), we use the 90th percentile cut-off. Regressions use the matched sample including the control variables of lagged *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Appendix Chapter A lists the detailed definitions and calculations of all variables in the regression. The unit of observation is fund-month. We cluster standard errors by fund and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6) TED	(7) Libor	(8) MOVE	(9) P90 VIX
Alternative	-0.2480	-0.1044	-0.5181	-0.8621*	-0.5013	-0.6912	-0.6779	-0.7053	-0.6494
	(0.5619)	(0.6332)	(0.5482)	(0.5157)	(0.5298)	(0.5189)	(0.5113)	(0.5770)	(0.5529)
Alternative x Stress	1.1808**	1.3295**	1.1951**	1.2982**	1.4562***	1.7446***	1.9577***	1.2229*	1.6387**
	(0.5446)	(0.5744)	(0.5319)	(0.5489)	(0.5652)	(0.6415)	(0.7220)	(0.6540)	(0.7516)
Stress	-1.2712***	-1.3592***	-1.2335***	-1.3075***	-1.4066***	-1.0096**	-1.0631*	-1.0970**	-1.1284*
	(0.3809)	(0.3795)	(0.3565)	(0.3884)	(0.3786)	(0.5134)	(0.5629)	(0.4866)	(0.5832)
Front load					0.1267				
					(0.0910)				
Rear load					0.9808				
					(0.6451)				
Observations	9,670	9,670	9,510	9,670	9,329	9,670	9,670	9,670	9,670
R-squared	0.035	0.030	0.028	0.040	0.029	0.026	0.026	0.025	0.026
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Region of Sale FE	Y								
Domicile FE		Y							
Loads					Y				
Global Category FE				Y					
Investment Area FE			Y						

Table C.16: *Summary information on switching funds.* Panel A shows the frequency table of switch dates funds which switch from being a traditionally priced fund to a fund with an alternative pricing rule. Panel B reports the differences in fund characteristics between switchers and their matched pairs during the event period from -24 months before to 24 months after the switch. Matching is performed on the last (monthly) observation before the switch occurs. We describe the matching algorithm in the text. *Treated* is an indicator variable that equals one for switching funds; *Post* is an indicator variable that equals one for the period after the switch. Columns (1) to (7) show results for *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, *Inst*, and *N of Inv*, respectively. Variable definitions are in Appendix Chapter A. * **, *** indicate 10%, 5%, and 1% level of significance, respectively.

Panel A. Dates of Switch

Switch Date	Freq.	Percent
2006-11	8	23.53
2007-10	3	8.82
2007-12	5	14.71
2010-11	2	5.88
2011-01	1	2.94
2011-03	2	5.88
2012-04	3	8.82
2012-05	6	17.65
2015-02	3	8.82
2016-01	1	2.94
Total	34	100.00

Panel B. Fund Characteristics during the Event Period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Alpha	Size	Age	Expense	Inst	Illiquidity	# Inv
Post	-0.0530	0.7280*	0.3698***	0.1061	-8.4894	0.0011	1.0645***
	(0.0759)	(0.4121)	(0.1133)	(0.0896)	(5.4755)	(0.0010)	(0.3967)
Treated	-0.0120	1.0215	0.2032	-0.1291	-14.3628	0.0022	0.6520
	(0.1508)	(0.6647)	(0.2412)	(0.1835)	(15.6695)	(0.0022)	(0.9748)
Post x Treated	-0.0638	-0.6143	-0.0418	-0.0856	6.2260	-0.0025	-0.6686
	(0.1463)	(0.4420)	(0.1258)	(0.0923)	(6.1336)	(0.0018)	(0.4258)
Constant	0.3595***	18.3541***	1.4856***	0.7350***	57.5929***	0.0083***	3.7339***
	(0.0760)	(0.6199)	(0.1953)	(0.1509)	(13.3108)	(0.0014)	(0.8401)
Observations	1,201	1,628	1,628	1,321	1,628	1,345	1,606
R-squared	0.009	0.059	0.085	0.041	0.023	0.010	0.026

Table C.17: *End-investor flows during market stress for switchers and their matched funds: Robustness tests.* This table shows the results of the robustness tests for the analysis reported in Table C.6. Event period is [-24, 24] months. Matching algorithm is described in the text. Unit of observation is investor-month. The analysis uses switchers and their matched funds. In column (1), the dependent variable is *Flow EndInv*, which is the percentage monthly change in each investor's holding (in number of shares). Column (1) is the same as column (3) of Table C.6 except that we choose the control funds from the same fund family and include investor×time fixed effects. Column (2) is the same as column (3) of Table C.6 except that we use *GBP Flow EndInv*, which is the GBP monthly change in each investor's holding (in thousands). *Treated* is an indicator variable that equals one for switching funds; *Stress* is an indicator variable that equals one if monthly VIX is above the 75th percentile of the sample; *Post* is an indicator variable that equals one for the period after the switch. Control variables include lagged values (previous month-end) of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Variable definitions are available in Appendix Chapter A. We cluster standard errors by investor and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1)	(2)
Stress x Treated x Post	0.3700**	20.2661**
	(0.1620)	(10.0143)
Stress x Post	-0.1614	-23.2665*
	(0.1627)	(13.5035)
Stress x Treated	-0.5964***	-11.3364
	(0.0870)	(11.0687)
Treated x Post	-0.3674***	-13.4425*
	(0.0978)	(7.2638)
Post	0.1628*	11.0915**
	(0.0952)	(4.9542)
Treated	-0.1211	-
	(0.1307)	
Stress	-	-5.3572
		(5.3542)
Alpha	-0.0758***	8.2812
	(0.0201)	(9.8101)
Size	0.2505***	-16.8371**
	(0.0391)	(7.4094)
Age	-0.7534***	-233.7087***
	(0.0803)	(57.6130)
Expense	-0.8612***	-94.3481
	(0.1518)	(60.6431)
Inst	0.0011	0.3571
	(0.0017)	(1.3463)
Illiquidity	8.1028***	535.5695
	(0.9325)	(578.2495)
Observations	272,601	402,843
R-squared	0.130	0.164

Table C.18: *Volatility of end-investor flows.* The sample includes investors in funds that changed their pricing rules (switchers) along with investors in the control group of no-switchers. Dependent variable is the volatility of *Flow EndInv*, defined as the percentage monthly change in each investor's holding, in number of shares. *Treated* is an indicator variable that equals one for switching funds and zero for the matched sample; *Post* is an indicator variable that equals one for the period after the switch. The event period is [-24, 24] months around the pricing change. Matching algorithm is described in the text. Control variables include lagged values (previous month-end) of *Alpha*, *Size*, *Age*, *Expense*, *Illiquidity*, and *Inst*. Variable definitions are available in Appendix Chapter A. We also include investor fixed effects. We cluster standard errors by investor and month. *, **, *** indicate 10%, 5%, and 1% level of significance, respectively.

VARIABLES	(1)
Treated x Post	-0.2121*
	(0.1119)
Post	0.2598**
	(0.1136)
Alpha	0.1491**
	(0.0658)
Size	-0.1124
	(0.1057)
Age	-0.8237*
	(0.4709)
Expense	-0.0645
	(0.3121)
Illiquidity	-0.1686
	(3.6408)
Inst	0.0098
	(0.0111)
Observations	15,824
R-squared	0.778
Investor FE	Y
Controls	Y

C.3 Figures

Figure C.1: *Daily VIX during the sample period.* The figure shows the daily (end-of-day day) values of Chicago Board Options Exchange Volatility Index (VIX) during our sample period, which is from January 2006 to December 2016. Vertical dashed lines indicate a number of important events. *Lehman* marks the bankruptcy of Lehman Brothers on September 15 2008; *Greek bailout* marks the launch of the bailout loan to Greece on 2 May 2010; *U.S. AA+* marks the downgrade of U.S. sovereign debt by S&P on 5 August 2011; *Draghi* marks the 26 July 2012 when Mario Draghi announced that the ECB is ready to do 'whatever it takes' to preserve the Euro; *TT* marks the beginning of the bond market crisis called 'Taper Tantrum' on 22 May 2013, and *ECB QE* marks the 10 March 2016 when the ECB increased its monthly bond purchases to €80bn and started including corporate bonds.

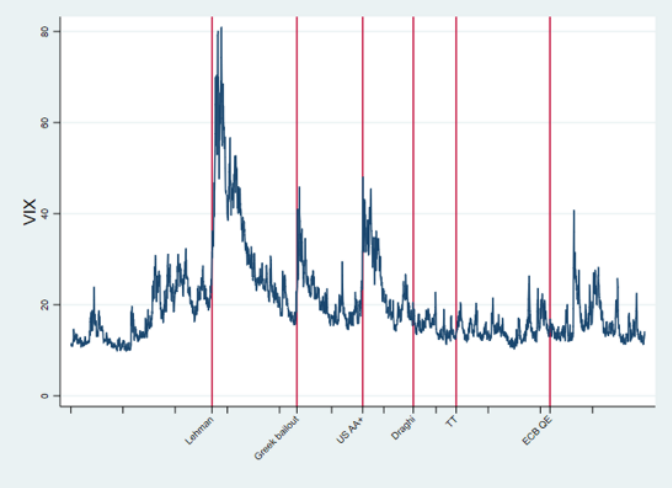


Figure C.2: *Dilution adjustment factor.* A fund's dilution adjustment factor, *Adjustment Factor*, is the factor by which the fund NAV is adjusted on a given day. It equals the *absolute value* of swing factor for swing funds; for dual funds, it equals the half spread of the difference in dual funds' bid and ask prices, $0.5 * (ask - bid)/mid$. Daily fund *Illiquidity* is the daily value-weighted average of bid-ask spreads of fund's assets. Vertical dashed lines indicate salient macroeconomic events described in Figure C.1.

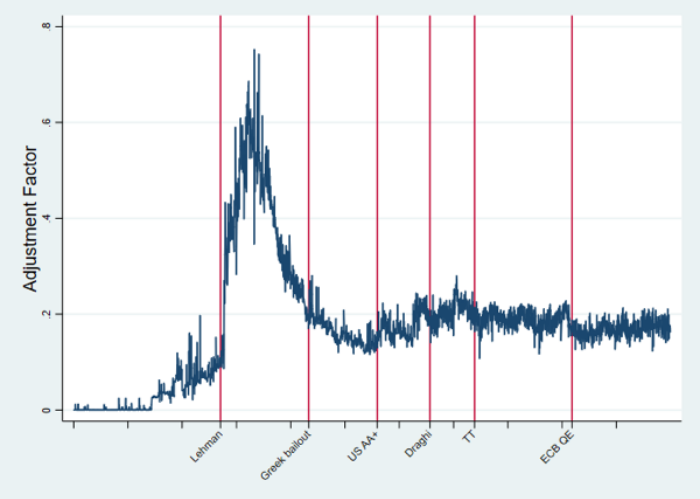


Figure C.3: *Fund characteristics before and after the switching event.* Figures below show the mean fund characteristics for switchers and their matched funds over the event period [-24 months, 24 months]. Blue lines represent mean values for treated funds (switchers); red lines represent mean values for control funds. Figures show *Alpha*, *Size*, *Age*, *Expense*, *Illiq*, *Inst*, and *N of Inv.* Variable definitions are available in Appendix Chapter A.

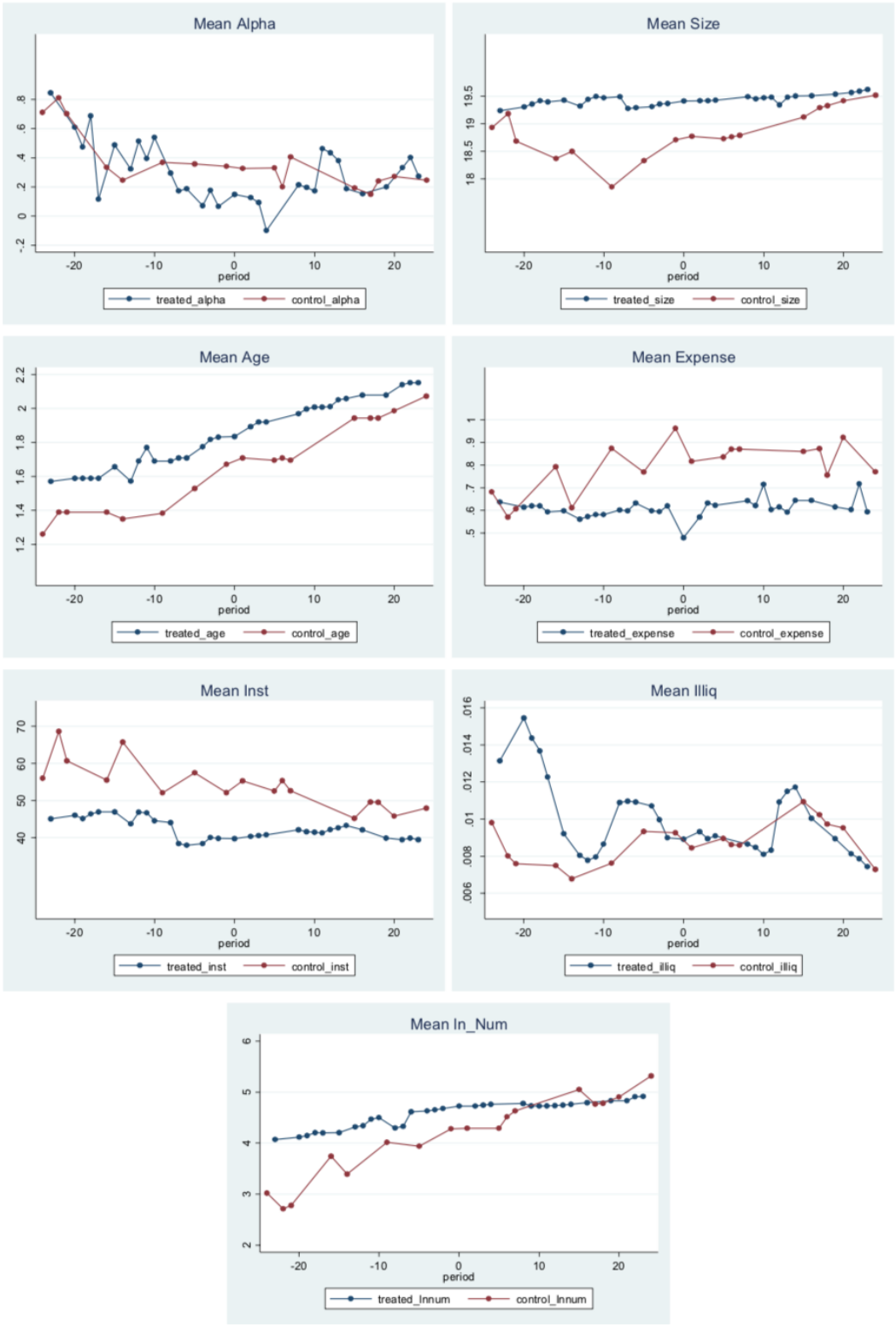
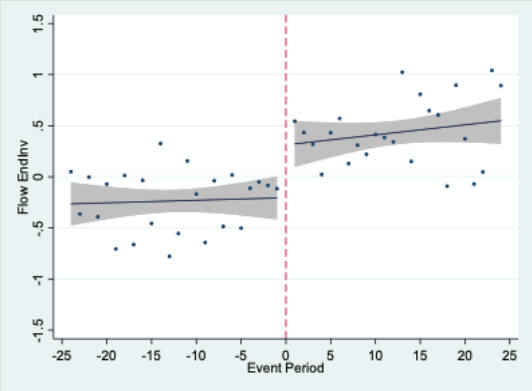
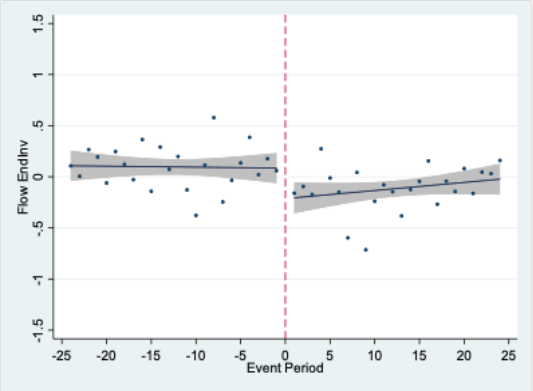


Figure C.4: *End investor fund flows before and after the switching event.*The graphs show the average difference in end investor flows, *Flow EndInv*, between switchers (treated) and their matched funds (control) after controlling for end-investor fixed effects. Differences are shown by event period over the event period, [-24 months, 24 months]. Panel A presents the plot for stress periods, and Panel B presents it for periods outside market stress. Figures include linear plots with 90% confidence intervals.

A. During Stress



B. Outside Stress

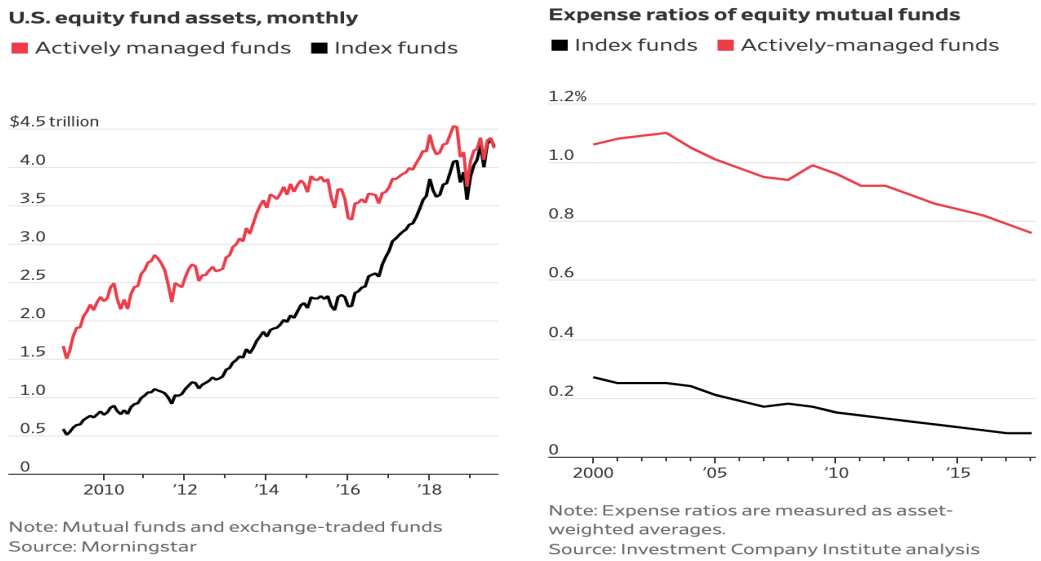


Appendix D

Appendix: The Drivers and Inhibitors of Factor Investing

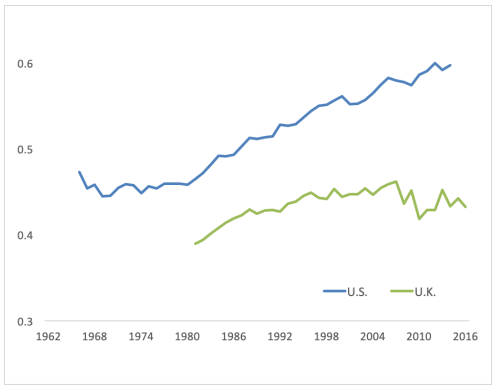
Abstract: This chapter contains the figures and data evidence (Section D.1), asset orthogonalization (Section D.2), proofs of the results (Section D.3), model framework and implications with heterogeneous households (Section D.4), and robustness in rich asset market and discussions of fund alpha (Section D.5).

D.1 Data Evidence

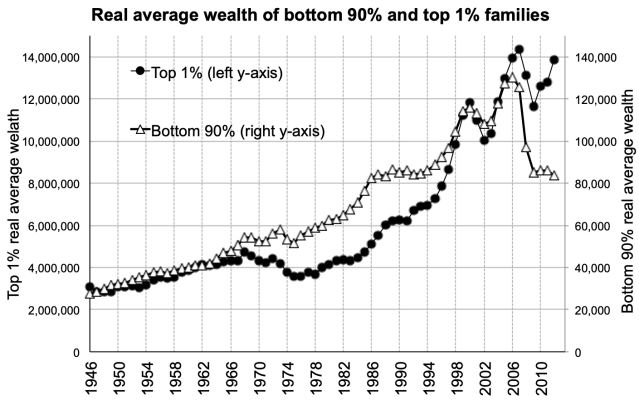


A. Capital B. Fees

Figure D.1: *U.S. delegtaion industry segmentation (Panel A) and delegation fees (Panel B).* In August 2019, the net asset value in indexing funds ($4.27 trillion), for the first time, surpasses their actively stock-selection rivals ($4.25 trillion). Meanwhile, the delegation fees for both types of funds decrease by roughly 0.3% for the past two decades.

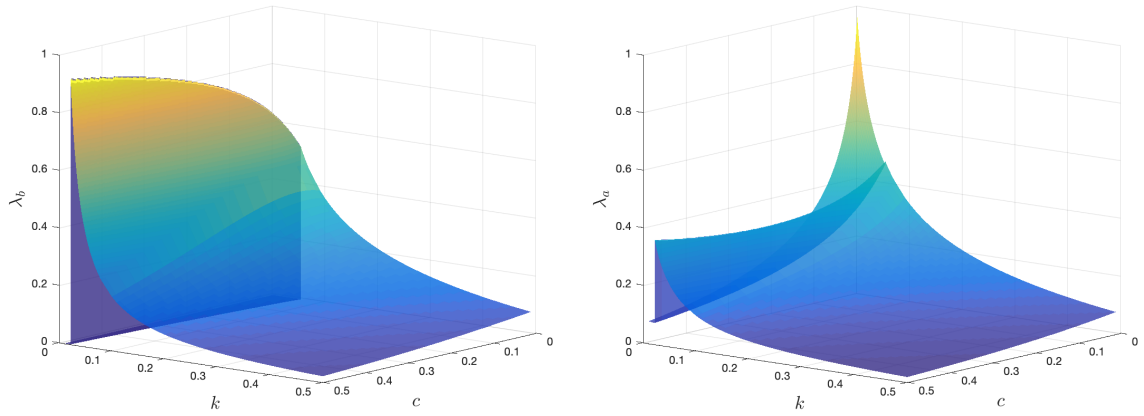


A. Gini coefficient



B. Wealth of bottom and top U.S. families

Figure D.2: *Time series evolution of wealth inequality.* Panel A plots the Gini coefficient of U.S. and U.K., 1962-2016, with data from World Bank. Panel B depicts the average real wealth of bottom 90% families (right y-axis) and top 1% families (left y-axis) in the U.S. from 1946 to 2012, with wealth expressed in constant 2010 U.S. dollars using the GDP deflator (details in Saez and Zucman (2016)).



A. Demand for factor investing

B. Demand for fundamental investing

Figure D.3: *Equilibrium capital allocation w.r.t. information costs.* As shown in Section 4.3, the equilibrium capital allocations of households are determined by a bunch of exogenous variables, $\{k, c, \gamma, \sigma_{q_2}^2, \sigma_e^2, \sigma_\varepsilon^2\}$. Varying the information cost pair $(k, c) \in (0, 0.5] \times (0, 0.5]$, the figure plots the equilibrium demand for factor investing λ_b (Panel A) and fundamental investing λ_a (Panel B). For the other parameters, take $\gamma = 3$, $\sigma_{q_I}^2 = \sigma_{q_2}^2 = \sigma_e^2 = \sigma_s^2 = 2\,\sigma_\varepsilon^2 = 0.2$.

D.2 Asset Orthogonalization

Because the risky underlying assets are correlated with each other both in terms of systematic factors and idiosyncratic factors, I apply the orthogonalization method to simplify the calculation of conditional expectation and covariance matrix. Note that there are multiple ways of orthogonalization yet in essence are identical. It is equivalent to finding an

orthogonal basis, among which the elements are mutually orthogonal, of the vector space spanned by $\boldsymbol{v}$. Similar to the synthetic assets in Bond and Garcıa (2017), the asset space is re-spanned to include one market portfolio, with payoff S, and $l-1$ stock-picking assets. I apply the standard Gram-Schmidt process to find the rest orthogonal basis of the underlying asset space, given in Lemma D.2.1.

Lemma D.2.1 (**Orthogonal basis**) *One l-dimensional orthogonal basis for the asset space is*

$$\begin{aligned}
\tilde{v}_1 &= S, \\
\tilde{v}_2 &= \mathcal{E}_2, \\
\tilde{v}_3 &= \mathcal{E}_3 - \frac{\langle \tilde{v}_2, \mathcal{E}_3 \rangle}{\langle \tilde{v}_2, \tilde{v}_2 \rangle} \tilde{v}_2 = \mathcal{E}_3 - \rho_{23}\, \mathcal{E}_2, \\
&\dots \\
\tilde{v}_l &= \mathcal{E}_l - \sum_{i=1}^{l-1} \frac{\langle \tilde{v}_i, \mathcal{E}_l \rangle}{\langle \tilde{v}_i, \tilde{v}_i \rangle} \tilde{v}_i,
\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the inner product operator.[1]

Taking $l = 3$ as an example, I show how an orthogonal basis works. The correlation between $\mathcal{E}_2$ and $\mathcal{E}_3$ is $\rho_{23} \coloneqq \frac{b_1^2 - b_2^2 - b_3^2}{2\, b_2\, b_3}$. Then a portfolio holding of $\boldsymbol{x} = (x_1, x_2, x_3)^\intercal$ with asset payoff $(v_1, v_2, v_3)^\intercal$ is equivalent to a portfolio holding of $\tilde{\boldsymbol{x}} = (x_1\, \beta_1 + x_2\, \beta_2 + x_3\, \beta_3, x_2 - x_1\, b_2/b_1 + \rho_{23}\, (x_3 - x_1\, b_3/b_1), x_3 - x_1\, b_3/b_1)^\intercal$ with asset payoff $(S, \mathcal{E}_2, \mathcal{E}_3 - \rho_{23}\, \mathcal{E}_2)^\intercal$. Intuitively, in the new asset space there are three orthogonal assets—the market index and two independent stock-picking assets. The independency greatly simplifies the calculation of portfolio choice, with the detailed results of conditional expectation and variance in the next section.

I specify the noisy supply vector of the re-spanned payoff space as follows. Denote the supply of the orthogonal base $(\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_l)$ in Lemma D.2.1 by $\tilde{\boldsymbol{q}} = (q_I, q_2, ..., q_l)$, normally distributed and independent element-wise, with the random variables satisfying $q_I \overset{d}{\sim} \mathcal{N}(\overline{q}_I, \sigma_{q_I}^2)$, and $q_i \overset{d}{\sim} \mathcal{N}(\overline{q}_i, \sigma_{q_i}^2)$ for $i \geq 2$. Assume the original and orthogonal asset spaces are linked through a nonsingular transforming matrix Γ, i.e., $\boldsymbol{v} = \Gamma\, \tilde{\boldsymbol{v}}$. Then the strategy and supply vector satisfy $\boldsymbol{x} = \Gamma\, \tilde{\boldsymbol{x}}$ and $\boldsymbol{q} = \Gamma\, \tilde{\boldsymbol{q}}$.

[1] For the detailed algorithm of Gram-Schmidt process and inner product, see, for instance, Cheney and Kincaid (2009).

D.3 Proofs

D.3.1 Equilibrium price

Proof of Proposition 4.2.1 In choosing the coefficients, I use subscript 'b' for the sensitivity of factor investors' signal towards price, subscript 'a' for the sensitivity of fundamental investors' signals, and subscript 'q' for the sensitivity of asset supply. I conjecture the linear forms (4.2.1) for asset prices, as functions of signals and supplies, i.e.,

$$\begin{aligned}
p_I &= \theta_0 + \theta_b \left[(s - \overline{s}) - \theta_q (q_I - \overline{q}_I) \right] = \theta_0 + \theta_b \hat{p_I}, \\
\tilde{p}_2 &= p_2 - \beta_2 p_I = \phi_a \left[(e_2 - 0) - \phi_q (q_2 - \overline{q}_2) \right], \\
\tilde{p}_3 &= p_3 - \beta_3 p_I = \psi_a \left[(e_3 - 0) - \psi_q \left((q_3 - \overline{q}_3) + \psi_\rho (q_2 - \overline{q}_2) \right) \right],
\end{aligned}$$

with the rest p_i following from orthogonalization. Note that in the assumed form I introduce an intermediate "auxiliary" price vector, $\hat{p}_I$, for computational convenience of conditional mean and variance. It contains the same information content as p_I. Before figuring out the parameters, I first determine the portfolio holding strategies for different investors. According to the facts from Grossman and Stiglitz (1980), the conditional mean and variance of two jointly normally distributed variables, X and Y are,

$$\begin{aligned}
\mathbb{E}[X|Y=y] &= \mathbb{E}[X] + \frac{Cov(X,Y)}{Var(Y)} (\, y - \mathbb{E}[Y] \,), \\
Var(X|Y=y) &= Var(X) - \frac{[\, Cov(X,Y) \,]^2}{Var(Y)}.
\end{aligned} \tag{D.3.1}$$

(i) Factor investors and closet indexers:
They observe an imperfect signal of market factor and invest in the market portfolio. By calculation we have

$$\mathbb{E}[S|\mathcal{I}_b] = \mathbb{E}[S|s] = \overline{s} + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_\varepsilon^2} (s - \overline{s}) = \overline{s} + \beta_{v,s} (s - \overline{s}),$$

and

$$Var(S|\mathcal{I}_b) = Var(S|s) = \frac{\sigma_s^2 \sigma_\varepsilon^2}{\sigma_s^2 + \sigma_\varepsilon^2} = \beta_{v,s} \sigma_\varepsilon^2.$$

Therefore, the strategy of factor investors (including closet indexers), decomposed using the orthogonal asset space, is

$$\tilde{\boldsymbol{x}}_b = \left(\frac{\mathbb{E}[S|s] - p_I}{\gamma \, Var(S|s)}, \, 0, ..., \, 0 \right)^\intercal. \tag{D.3.2}$$

(ii) Direct investors:

The households who choose to invest directly in the asset market, observe nothing but the public listed prices. By calculation we have

$$\mathbb{E}[S|\mathcal{I}_h] = \mathbb{E}[S|p_I] = \mathbb{E}[S|\hat{p}_I] = \overline{s} + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_\varepsilon^2 + \theta_q^2\,\sigma_{q_I}^2}\,\hat{p}_I = \overline{s} + \beta_{v,p_I}\,\hat{p}_I,$$

and

$$Var(S|\mathcal{I}_h) = Var(S|p_I) = (1 - \beta_{v,p_I})\,\sigma_s^2.$$

Therefore, the strategy of direct investors, decomposed using the new orthogonal asset space, is

$$\tilde{\boldsymbol{x}}_h = \left(\frac{\mathbb{E}[S|p_I] - p_I}{\gamma\, Var(S|p_I)},\, 0, ...,\, 0\right)^\intercal. \tag{D.3.3}$$

(iii) Fundamental investors:

The skilled active managers observe the signals of both systematic and idiosyncratic loadings. Similarly, I define

$$\beta_{v,e} = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_\varepsilon^2}.$$

Re-spanning the payoff space guarantees independent assets and a diagonal variance-covariance matrix:

$$\mathbb{E}\left[\begin{array}{c} S \\ \mathcal{E}_2 \\ \mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2 \\ \cdots \end{array}\;\middle|\;\mathcal{I}_a\right] = \begin{bmatrix} \mathbb{E}[S|s] \\ \mathbb{E}[\mathcal{E}_2|e_2] \\ \mathbb{E}[\mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2|e_2, e_3] \\ \cdots \end{bmatrix} = \begin{bmatrix} \overline{s} + \beta_{v,s}\,(s - \overline{s}) \\ \beta_{v,e}\,e_2 \\ \beta_{v,e}\,(e_3 - \rho_{23}\,e_2) \\ \cdots \end{bmatrix},$$

and

$$Var\left(\begin{array}{c} S \\ \mathcal{E}_2 \\ \mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2 \\ \cdots \end{array}\;\middle|\;\mathcal{I}_a\right) = \begin{bmatrix} Var(S|s) & & & \\ & Var(\mathcal{E}_2|e_2) & & \\ & & Var(\mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2|e_2, e_3) & \\ & & & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} \beta_{v,s}\,\sigma_\varepsilon^2 & & & \\ & \beta_{v,e}\,\sigma_\varepsilon^2 & & \\ & & (1 + \rho_{23}^2)\,\beta_{v,e}\,\sigma_\varepsilon^2 & \\ & & & \cdots \end{bmatrix}.$$

Note that conditional variance sits on the independency of noise terms,

$$Cov(\mathcal{E}_2, \mathcal{E}_3|e_2, e_3) = Cov(\varepsilon_2, \varepsilon_3|e_2, e_3) = 0.$$

Therefore, the strategy of fundamental investors, decomposed using the orthogonal asset space, is

$$\tilde{\boldsymbol{x}}_a = \left(\frac{\mathbb{E}[S|s] - p_I}{\gamma\, Var(S|s)}, \frac{\mathbb{E}[\mathcal{E}_2|e_2] - \tilde{p}_2}{\gamma\, Var(\mathcal{E}_2|e_2)}, \frac{\mathbb{E}[\mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2|e_2, e_3] - \tilde{p}_3 + \rho_{23}\,\tilde{p}_2}{\gamma\, Var(\mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2|e_2, e_3)}, ... \right)^{\intercal}. \tag{D.3.4}$$

Now I determine the coefficients for equilibrium prices. Apply the market clearing condition in the orthogonalized asset space (in essence identical to condition (4.2.5)), as follows,

$$\lambda_h\, \tilde{\boldsymbol{x}}_h + \big(\lambda_b + (1-\mu)\,(1-\xi)\,\lambda_a \big)\, \tilde{\boldsymbol{x}}_b + \big(\xi + \mu\,(1-\xi) \big)\, \lambda_a\, \tilde{\boldsymbol{x}}_a = \tilde{\boldsymbol{q}}, \tag{D.3.5}$$

where $\tilde{\boldsymbol{q}}$ is the noisy supply vector of the index and stock-picking assets, introduced in Appendix Section D.2. Plugging in the demands $\tilde{\boldsymbol{x}}_j$ ($j \in \{h, p, a\}$) developed above, the equations turn out to be linear in the random signals and supplies. As they have to hold for any values of q_I, q_i and s, e_i, the aggregate coefficients on these variables have to equal zero, as well as the constant term.

Solving the market clearing equation for market portfolio, with payoff S, yields

$$\begin{aligned}
\theta_q &= \frac{\gamma\, \sigma_\varepsilon^2}{\lambda_b + \lambda_a}, \\
\theta_b &= \frac{(\lambda_b + \lambda_a)\, \beta_{v,s} + \lambda_h\, \frac{Var(S|s)}{Var(S|p_I)}\, \beta_{v,p_I}}{(\lambda_b + \lambda_a) + \lambda_h\, \frac{Var(S|s)}{Var(S|p_I)}}, \\
\theta_0 &= \overline{s} - \frac{\overline{q}_I\, \gamma\, Var(S|s)}{(\lambda_b + \lambda_a) + \lambda_h\, \frac{Var(S|s)}{Var(S|p_I)}}.
\end{aligned}$$

Solving the market clearing equation for the first stock-picking asset, with payoff $\mathcal{E}_2$, yields

$$\begin{aligned}
\phi_q &= \frac{\gamma\, \sigma_\varepsilon^2}{\big(\xi + \mu\,(1-\xi) \big)\, \lambda_a}, \\
\phi_a &= \beta_{v,e}, \\
\overline{q}_2 &= 0.
\end{aligned}$$

Solving the market clearing equation for the second stock-picking asset, with payoff $\mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2$, yields

$$\begin{aligned}
\psi_q &= (1 + \rho_{23}^2)\, \phi_q, \\
\psi_a &= \phi_a, \\
\psi_\rho &= \frac{\rho_{23}}{1 + \rho_{23}^2}, \\
\overline{q}_3 &= 0.
\end{aligned}$$

Similar analysis can be applied to the rest stock-picking assets. It turns out that $\overline{q}_2 = \overline{q}_3 = ... = \overline{q}_l = 0$. Intuitively, this implies that the expected supply of $v_1, v_2, ..., v_l$ is proportional to their fractions in the index. Thus the weights of the stock components in index are proportional to the stocks' total market capitalization.

D.3.2 Price efficiency and relative utility

Landing on the results of Proposition 4.2.1 and definition (4.2.6), we have

$$\eta_I = -\frac{1}{2}\log\left(1 - \frac{\sigma_\varepsilon^2\,\sigma_s^2\,\sigma_{q_I}^2/(\sigma_s^2+\sigma_\varepsilon^2)}{(\lambda_b+\lambda_a)^2/\gamma^2 + \sigma_\varepsilon^2\,\sigma_{q_I}^2}\right),$$

where $\sigma_{q_I}^2$ is the supply uncertainty of market portfolio. One can see that η_I decreases with the amount of capital in fund delegation. Moreover, the price inefficiency does not depend on the fraction of skilled managers μ or informed households ξ. This is because the idiosyncratic information, in aggregation, does not affect the informativeness of the price of market portfolio.

Similar to the index price inefficiency (4.2.6), I define the price inefficiency of idiosyncratic attributes. Due to the symmetric property in choosing orthogonal basis, I take $\mathcal{E}_2$ as an example and perform necessary calculation to get,

$$\tilde{\eta}_2 \coloneqq -\frac{1}{2}\log\left(\frac{Var(\mathcal{E}_2|e_2)}{Var(\mathcal{E}_2|\tilde{p}_2)}\right) = -\frac{1}{2}\log\left(1 - \frac{\sigma_\varepsilon^2\,\sigma_e^2\,\sigma_{q_2}^2/(\sigma_e^2+\sigma_\varepsilon^2)}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2/\gamma^2 + \sigma_\varepsilon^2\,\sigma_{q_2}^2}\right) \in (0,\infty), \tag{D.3.6}$$

where $\sigma_{q_2}^2$ is the supply uncertainty of the stock-picking asset with payoff $\mathcal{E}_2$. This helps investigate the price inefficiency of the individual asset. For stock 2, its price inefficiency is defined to be and equal to

$$\eta_2 \coloneqq -\frac{1}{2}\log\left(\frac{Var(v_2|\mathcal{I}_a)}{Var(v_2|\boldsymbol{p})}\right) = -\frac{1}{2}\log\left(\frac{\beta_2^2\,\frac{\sigma_s^2}{\sigma_s^2+\sigma_\varepsilon^2} + \frac{\sigma_e^2}{\sigma_e^2+\sigma_\varepsilon^2}}{\beta_2^2\,\frac{\sigma_s^2}{\sigma_s^2+\sigma_\varepsilon^2}\,\frac{(\lambda_b+\lambda_a)^2+\gamma^2\,\sigma_\varepsilon^2\,\sigma_{q_I}^2}{(\lambda_b+\lambda_a)^2+\gamma^2\,\sigma_\varepsilon^4\,\sigma_{q_I}^2} + \frac{\sigma_e^2}{\sigma_e^2+\sigma_\varepsilon^2}\,\frac{(\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2+\gamma^2\,\sigma_\varepsilon^2\,\sigma_{q_2}^2}{(\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2+\gamma^2\,\sigma_\varepsilon^4\,\sigma_{q_2}^2}}\right).$$

Recall the notations $\beta_{v,s} = \frac{\sigma_s^2}{\sigma_s^2+\sigma_\varepsilon^2}$ and $\beta_{v,e} = \frac{\sigma_e^2}{\sigma_e^2+\sigma_\varepsilon^2}$. The three inefficiency expressions can be linked together in a following way,

$$e^{-2\,\eta_2} = \frac{Var(v_2|\mathcal{I}_a)}{Var(v_2|\boldsymbol{p})} = \frac{\beta_2\,Var(S|s) + Var(\mathcal{E}_2|e_2)}{\beta_2\,Var(S|p_I) + Var(\mathcal{E}_2|\tilde{p}_2)} = \frac{\beta_2\,\beta_{v,s} + \beta_{v,e}}{\beta_2\,\beta_{v,s}\,e^{2\,\eta_I} + \beta_{v,e}\,e^{2\,\tilde{\eta}_2}}. \tag{D.3.7}$$

The price efficiency can be linked to the relative utility generated by fund performance, in the following way. Take $l = 3$ as an example.

Lemma D.3.1 (**Relative utility**) *The excess performance of delegation over household uninformed investing, represented using the relative utility, can be linked to the price efficiency:*

$$\gamma\,(u_b - u_h) = \eta_I,$$

$$\gamma\,(u_a - u_h) = \eta_I + \frac{1}{2}\log\Big(1+\frac{\gamma^2\,\beta_{v,e}\,\sigma_\varepsilon^2\,\sigma_{q_2}^2}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2}\Big) + \frac{1}{2}\log\Big(\frac{1+\rho_{23}^2\,(2\beta_{v,e}-1)}{1+\rho_{23}^2} + \frac{\gamma^2\,(1+\rho_{23}^2)\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2}\,\sigma_{q_3}^2\Big).$$

Proof: First note that

$$e^{-\gamma\,u_j} = \mathbb{E}\left[\exp\Big\{-\frac{1}{2}\left(\mathbb{E}[\boldsymbol{v}|\mathcal{I}_j]-\boldsymbol{p}\right)^{\intercal}\boldsymbol{\Sigma}_j^{-1}\left(\mathbb{E}[\boldsymbol{v}|\mathcal{I}_j]-\boldsymbol{p}\right)\Big\}\right],$$

where $\boldsymbol{\Sigma}_j$ is the conditional variance-covariance matrix of $\boldsymbol{v}$, and $\mathbb{E}[\boldsymbol{v}|\mathcal{I}_j]$ is the conditional mean, for any investor type $j \in \{h,p,a\}$. After orthogonalization, we have for any type $j \in \{h,p\}$,

$$e^{-\gamma\,u_j} = \mathbb{E}\left[\exp\left\{-\frac{1}{2}\,\frac{(\mathbb{E}[S|\mathcal{I}_j]-p_I)^2}{Var(S|\mathcal{I}_j)}\right\}\right];$$

and for fundamental investors ($j=a$),

$$e^{-\gamma\,u_a} = \mathbb{E}\left[\exp\left\{-\frac{1}{2}\left(\frac{(\mathbb{E}[S|\mathcal{I}_a]-p_I)^2}{Var(S|\mathcal{I}_a)} + \frac{(\mathbb{E}[\mathcal{E}_2|\mathcal{I}_a]-\tilde{p}_2)^2}{Var(\mathcal{E}_2|\mathcal{I}_a)} + \frac{(\mathbb{E}[\mathcal{E}_3-\rho_{23}\,\mathcal{E}_2|\mathcal{I}_a]-\tilde{p}_3+\rho_{23}\,\tilde{p}_2)^2}{Var(\mathcal{E}_3-\rho_{23}\,\mathcal{E}_2|\mathcal{I}_a)}\right)\right\}\right].$$

To complete the proof, I need the following facts. First, for independently normally distributed variables $z_i \overset{d}{\sim} \mathcal{N}(\mu_i, V_i)$, $i=1,2,3$, we have

$$\mathbb{E}\left[\exp\Big\{-\frac{1}{2}\left(z_1^2+z_2^2+z_3^2\right)\Big\}\right] = \prod_{i=1}^{3}(1+V_i)^{-\frac{1}{2}}\,\exp\Big\{-\frac{1}{2}\,\frac{\mu_i^2}{1+V_i}\Big\},$$

which comes form the moment-generating function of multi-dimensional non-central chi-square distribution. Second, according to the facts (D.3.1) listed in the proof of Proposition 4.2.1, we have

$$Var(\,\mathbb{E}[X|Y]\,) = \frac{[\,Cov(X,Y)\,]^2}{Var(Y)} = Var(X) - Var(X|Y), \tag{D.3.8}$$

which in fact comes from the law of total variance: the unconditional variance of a posterior belief equals the variance of the prior minus the posterior variance.

Then, performing necessary calculations yields

$$u_a = \frac{1}{2\,\gamma}\Bigg[\log\Big(\frac{Var(S-p_I)}{Var(S|s)}\Big)+\log\Big(\frac{Var(\mathcal{E}_2-\tilde{p}_2)}{Var(\mathcal{E}_2|e_2)}\Big)+\log\Big(\frac{Var(\mathcal{E}_3-\rho_{23}\,\mathcal{E}_2-\tilde{p}_3+\rho_{23}\,\tilde{p}_2)}{Var(\mathcal{E}_3-\rho_{23}\,\mathcal{E}_2|e_2,e_3)}\Big) + \frac{(\overline{s}-\theta_0)^2}{Var(S-p_I)}\Bigg],$$

$$u_b = \frac{1}{2\,\gamma}\left[\log\Big(\frac{Var(S-p_I)}{Var(S|s)}\Big)+\frac{(\overline{s}-\theta_0)^2}{Var(S-p_I)}\right],$$

$$u_h = \frac{1}{2\,\gamma}\left[\log\Big(\frac{Var(S-p_I)}{Var(S|p_I)}\Big)+\frac{(\overline{s}-\theta_0)^2}{Var(S-p_I)}\right].$$

Note that the last term in the certainty-equivalent wealth, $\frac{1}{2\,\gamma}\frac{(\overline{s}-\theta_0)^2}{Var(S-p_I)}$, stands for the monetary utility of investing in index through unconditional mean-variance strategies.

Given the definition of index price inefficiency, equation (4.2.6), it is straightforward that

$$\gamma\,(u_b-u_h)=\eta_I.$$

Now I investigate the excess performance of active funds. Based on the results from Proposition 4.2.1, one can see that

$$\gamma\,(u_a-u_h)=\eta_I+\frac{1}{2}\,\log\Big(\frac{Var(\mathcal{E}_2-\tilde{p}_2)}{\beta_{v,e}\,\sigma_\varepsilon^2}\Big)+\frac{1}{2}\,\log\Big(\frac{Var(\mathcal{E}_3-\rho_{23}\,\mathcal{E}_2-\tilde{p}_3+\rho_{23}\,\tilde{p}_2)}{(1+\rho_{23}^2)\,\beta_{v,e}\,\sigma_\varepsilon^2}\Big). \tag{D.3.9}$$

Performing necessary calculation yields the unconditional variance of the stock-picking asset returns,

$$Var(\mathcal{E}_2-\tilde{p}_2)=(1-\phi_a)^2\,\sigma_e^2+\phi_a^2\,\sigma_\varepsilon^2+\phi_a^2\,\phi_q^2\,\sigma_{q_2}^2=\beta_{v,e}\,\sigma_\varepsilon^2\left(1+\frac{\gamma^2\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2}\,\sigma_{q_2}^2\right), \tag{D.3.10}$$

and

$$\begin{aligned}Var(\mathcal{E}_3-\rho_{23}\,\mathcal{E}_2-\tilde{p}_3+\rho_{23}\,\tilde{p}_2)&=(1-\rho_{23}^2)\,(1-\phi_a)^2\,\sigma_e^2+(1+\rho_{23}^2)\,\phi_a^2\,\sigma_\varepsilon^2+(1+\rho_{23}^2)^2\,\phi_a^2\,\phi_q^2\,\sigma_{q_3}^2\\&=\beta_{v,e}\,\sigma_\varepsilon^2\left(1+\rho_{23}^2\,(2\,\beta_{v,e}-1)+\frac{\gamma^2\,(1+\rho_{23}^2)^2\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2}\,\sigma_{q_3}^2\right).\end{aligned} \tag{D.3.11}$$

Thus equation (D.3.9) becomes

$$\begin{aligned}\gamma\,(u_a-u_h)=\eta_I+\frac{1}{2}\,\log\Big(1+\frac{\gamma^2\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2}\,\sigma_{q_2}^2\Big)+\\ \frac{1}{2}\,\log\Big(\frac{1+\rho_{23}^2\,(2\,\beta_{v,e}-1)}{1+\rho_{23}^2}+\frac{\gamma^2\,(1+\rho_{23}^2)\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2}\,\sigma_{q_3}^2\Big).\end{aligned}$$

□

Let us take a look at the extreme scenario where fund managers have no inside information, mathematically, $\sigma_\varepsilon^2 = +\infty$. By definition (4.2.6) it is not hard to figure out that $\eta_I = \tilde{\eta}_2 = 0$, in other words, all the available information is contained in the price. Obviously, factor investing can do no better than direct investing utility-wise, $u_b = u_h$. However, due to the investment restrictions that, both factor funds and households themselves can only access the market portfolio in aggregation, for "hedging" purpose the active managers could still earn some market share, even in the case where they hold no advanced information. This can be seen from the relative utility result in Lemma D.3.1.

The main interest lies in the non-extreme cases, where the signal is neither perfect nor completely useless, i.e., $\sigma_\varepsilon^2 \in (0, \infty)$. A natural concern is how does *information accuracy* affect price efficiency. Intuitively, a smaller σ_ε^2 means a more accurate signal. The effect of signal precision depends on how active the assets are traded. A relatively *liquid* market is featured by more frequent trading behavior, thus, in my model is characterized by a relatively larger $\sigma_{q_i}^2$ and $\sigma_{q_I}^2$. The relation between information accuracy and price efficiency is stated below.

Lemma D.3.2 (**Information accuracy**) *(i) If market is relatively liquid, i.e., $\sigma_s^2 \leq \theta_q^2\,\sigma_{q_I}^2$ and $\sigma_e^2 \leq \phi_q^2\,\sigma_{q_2}^2$, then an increase in information accuracy increases the price efficiency of index and stock-picking assets;*

(ii) If market is relatively illiquid, i.e., $\sigma_s^2 > \theta_q^2\,\sigma_{q_I}^2$ and $\sigma_e^2 > \phi_q^2\,\sigma_{q_2}^2$, then an increase in information accuracy weakly decreases the price efficiency of index and stock-picking assets.

Proof: Recall the expressions of η_I, $\tilde{\eta}_2$, and η_2. The higher the numbers, the less efficient the prices are. Now I investigate the relation between σ_ε^2 and them. For the market index, for instance, consider the following derivative

$$\mathcal{D}\Big[\frac{\sigma_\varepsilon^2\,\sigma_s^2\,\sigma_{q_I}^2/(\sigma_s^2+\sigma_\varepsilon^2)}{(\lambda_b+\lambda_a)^2/\gamma^2+\sigma_\varepsilon^2\,\sigma_{q_I}^2},\,\sigma_\varepsilon^2\Big].$$

Performing necessary calculation, it is positive if and only if

$$\sigma_s^2 > \theta_q^2\,\sigma_{q_I}^2. \tag{D.3.12}$$

That is to say, if condition (D.3.12) is satisfied, then an increase in σ_ε^2 increases η_I thus deteriorates the price efficiency. Performing similar analysis on $\tilde{\eta}_2$, we see the results in the lemma hold. □

Another controversial topic to practitioners and academia is that, how does the increase in factor investing affect the price informativeness of relatively liquid vs. illiquid stocks? A more actively traded stock, in my framework, is mapped into the one with a relatively larger $\sigma_{q_i}^2$. Then Lemma D.3.3 gives answer to the question.

Lemma D.3.3 (**Factor investing on liquid vs. illiquid assets**) *Holding the total fraction of wealth under delegation fixed,*

(i) The rising factor investing has larger impact on the price informativeness of illiquid assets than on the liquid ones;

(ii) If an asset is extremely illiquid in terms of its idiosyncratic part, then the price is almost perfectly efficient regardless of the change of capital allocation.

Proof: I compare two assets with different liquidity to show the results. Counting on definition (D.3.6), we have

$$\tilde{\eta}_2 \coloneqq -\frac{1}{2}\log\Big(\frac{Var(\mathcal{E}_2|e_2)}{Var(\mathcal{E}_2|\tilde{p}_2)}\Big), \quad \tilde{\eta}_3 \coloneqq -\frac{1}{2}\log\Big(\frac{Var(\mathcal{E}_3|e_3)}{Var(\mathcal{E}_3|\tilde{p}_3)}\Big).$$

To get the price efficiency of idiosyncratic loading $\mathcal{E}_3$, I simply resort to another orthogonal basis—$(S, \mathcal{E}_3, \mathcal{E}_2 - \rho_{23}\,\mathcal{E}_3)$, where ρ_{23} is still the correlation between $\mathcal{E}_2$ and $\mathcal{E}_3$. The second element of the new basis, exactly what is needed, is now isolated. I also link the two basis with a transforming matrix:

$$\begin{bmatrix} S \\ \mathcal{E}_3 \\ \mathcal{E}_2 - \rho_{23}\,\mathcal{E}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho_{23} & 1 \\ 0 & 1-\rho_{23}^2 & -\rho_{23} \end{bmatrix} \begin{bmatrix} S \\ \mathcal{E}_2 \\ \mathcal{E}_3 - \rho_{23}\,\mathcal{E}_2 \end{bmatrix}.$$

Thus the noise of the supply of new basis with re-spanned payoff, $(S, \mathcal{E}_3, \mathcal{E}_2 - \rho_{23}\,\mathcal{E}_3)^\intercal$, can be expressed using the old ones as $(\sigma_{q_I}^2, \rho_{23}^2\,\sigma_{q_2}^2 + \sigma_{q_3}^2, (1-\rho_{23}^2)^2\,\sigma_{q_2}^2 + \rho_{23}^2\,\sigma_{q_3}^2)^\intercal$.

Without loss of generality, assume $\mathcal{E}_2$ is more liquid than $\mathcal{E}_3$, i.e., $\sigma_{q_2}^2 > \rho_{23}^2\,\sigma_{q_2}^2 + \sigma_{q_3}^2$. Following the above results, the price inefficiency of the two assets are

$$\begin{cases} \tilde{\eta}_2 &= -\dfrac{1}{2}\log\left(1 - \dfrac{\sigma_\varepsilon^2\,\sigma_e^2/(\sigma_e^2+\sigma_\varepsilon^2)}{(\,\xi + \mu\,(1-\xi)\,)^2\,\lambda_a^2/\gamma^2\,\sigma_{q_2}^2 + \sigma_\varepsilon^2}\right), \\ \tilde{\eta}_3 &= -\dfrac{1}{2}\log\left(1 - \dfrac{\sigma_\varepsilon^2\,\sigma_e^2/(\sigma_e^2+\sigma_\varepsilon^2)}{(\,\xi + \mu\,(1-\xi)\,)^2\,\lambda_a^2/\gamma^2\,(\rho_{23}^2\,\sigma_{q_2}^2 + \sigma_{q_3}^2) + \sigma_\varepsilon^2}\right). \end{cases}$$

Comparing the coefficients before λ_a in the expressions of price inefficiency yields the results.

If an asset, say $\mathcal{E}_2$, is extremely illiquid, i.e., $\sigma_{q_2}^2 \to 0$. Then from the expression one can conclude that $\tilde{\eta}_2 \to 0$, intuitively, the price is always perfectly informative. □

D.3.3 Equilibria and stylized facts

Proof of Lemma 4.3.1 The results follow directly from the relative utility outcome in Lemma D.3.1. Also in the proof of interior equilibrium I will show that to guarantee the inflows of active industry, μ cannot be too small. As long as $\mu > 0$, in any equilibrium, before fees we have

$$\mu\, u_a + (1-\mu)\, u_b > u_b.$$

Proof of Proposition 4.3.1 I mainly use condition (4.3.8) to determine λ_b. Rearranging the equation yields

$$u_a - \alpha_a - c = u_a - (u_b - u_h + k) - c = u_h,$$

which, after simplification, implies

$$\gamma\,(u_a - u_b) = \gamma\,(k + c). \tag{D.3.13}$$

Counting on the relative utility result from Lemma D.3.1, with two risky assets (one market portfolio and one stock-picking asset) we have

$$\gamma\,(u_a - u_b) = \frac{1}{2}\,\log\Big(1 + \frac{\gamma^2\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi + \mu\,(1-\xi)\,)^2\,\lambda_a^2}\,\sigma_{q_2}^2\Big). \tag{D.3.14}$$

Note that equation (4.3.13), (4.3.16), and (4.3.17) imply the following relationship:

$$\big(\,\xi + \mu\,(1-\xi)\,\big)\,\lambda_a = \frac{k\,\gamma\,\mu + \eta_I\,\mu}{k\,\gamma\,\mu + \eta_I}\,\lambda_a = \mu\,\lambda_b.$$

Counting on this result and equating expression (D.3.13) and (D.3.14), one can solve for the equilibrium fraction of capital in factor investing, as follows,

$$\lambda_b = \frac{\gamma\,\sigma_\varepsilon\,\sigma_{q_2}}{\mu}\,\sqrt{\frac{\beta_{v,e}}{e^{2\,\gamma\,(k+c)} - 1}}. \tag{D.3.15}$$

Because μ is uniquely determined given information costs k and c, λ_b is uniquely determined.

Then let us take a look at λ_a. The total capital in funds is related to the index price efficiency. Thus from definition (4.2.6), one get

$$\lambda_a + \lambda_b = \gamma\,\sigma_\varepsilon\,\sigma_{q_I}\sqrt{\frac{\beta_{v,s}}{1-e^{-2\,\eta_I}}-1}.$$

Given the ratio between demand for factor and fundamental investing (4.3.17), the following expression holds:

$$\frac{\lambda_a+\lambda_b}{\lambda_b} = \frac{(\mu+1)\,k\,\gamma+2\,\eta_I}{k\,\gamma+\eta_I}.$$

As we already know the equilibrium demand for factor investing λ_b, I first combine the results to solve η_I:

$$\frac{\lambda_a+\lambda_b}{\mu\,\lambda_b} = \frac{(\mu+1)\,k\,\gamma+2\,\eta_I}{\mu\,(k\,\gamma+\eta_I)} = \frac{\sigma_{q_I}}{\sigma_{q_2}}\,\frac{\sqrt{\frac{\beta_{v,s}}{1-e^{-2\,\eta_I}}-1}}{\sqrt{\frac{\beta_{v,e}}{e^{2\,\gamma\,(k+c)}-1}}}. \tag{D.3.16}$$

Then equilibrium active wealth λ_a follows.

In equation (D.3.16), the LHS is monotonically increasing in η_I, and the RHS is monotonically decreasing in η_I. In addition, both sides vary within a positive support region. Hence, once there exists an η_I equating the two sides, it must be unique. One can regard it as the intersection point of two functions of η_I. The case where such η_I does not exist would be discussed in the boundary equilibria cases.

Proof of Proposition 4.3.2 Given $\lambda_b + \lambda_a = 1$, the model directly determines the efficiency of index price:

$$\eta_I = -\frac{1}{2}\,\log\left(1-\frac{\gamma^2\,\sigma_\varepsilon^2\,\sigma_s^2\,\sigma_{q_I}^2/(\sigma_s^2+\sigma_\varepsilon^2)}{1+\gamma^2\,\sigma_\varepsilon^2\,\sigma_{q_I}^2}\right).$$

From the ratio

$$\frac{\lambda_b}{\lambda_a} = \frac{\lambda_b}{1-\lambda_b} = \frac{\alpha_b+k}{\alpha_b+\mu\,k},$$

the factor fund fees α_b is uniquely determined, then by relation (4.3.11) α_a as well:

$$\alpha_b = \frac{k\big[1-(1+\mu)\,\lambda_b\big]}{2\,\lambda_b-1}, \qquad \alpha_a = \frac{k\,(1-\mu)\,\lambda_b}{2\,\lambda_b-1}.$$

From condition (4.3.14) and (4.3.15) I can express some key parameters as functions of α_b:

$$\xi = \frac{\mu\,k}{\mu\,k+\alpha_b}, \qquad \mu = \frac{e^{\gamma\,k}-1}{e^{\gamma\,k}-e^{-\gamma\,c}} \in \big[1-e^{-\gamma\,k},1\big].$$

One can determine demand for factor investing, λ_b, similarly using relative utility results as in the proof of Proposition 4.3.1. Then $\lambda_a = 1 - \lambda_b$. Up till now the equilibrium has been solved analytically.

Now I take a look at how the active delegation fees change when demand for factor investing λ_b varies. Rearranging the expression of α_a in the proposition yields

$$\alpha_a = \frac{k\,(1-\mu)}{2 - 1/\lambda_b}.$$

Thus an increase in λ_b, possibly brought up by the exogenous costs or supply shocks, most of the time is accompanied by active funds lowering their fees α_a on average.

Proof of Lemma 4.3.2 For the delegation dominating equilibrium to hold, we must have $\alpha_b \geq 0$ and $\xi \leq 1$. Thus the following conditions must hold,

$$\begin{cases} 1 - \lambda_b - \mu\,\lambda_b \geq 0, \\ 2\,\lambda_b - 1 > 0, \end{cases}$$

which yields the necessary conditions in the lemma.

Proof of Proposition 4.3.3 The coexistence of factor and active fund managers, condition (4.3.15), together with $\lambda_b = 0$, implies that all the households would like to incur a cost to become informed towards manager quality, i.e., $\xi = 1$. Thus applying condition (4.3.14) we get

$$\alpha_a = k, \qquad \alpha_b = \alpha_a - k = 0,$$

in other words, the funds are driven to no profit boundary. With zero delegation fees, one can see that $u_b = u_h$, in which case the price of market portfolio is fully informative ($\eta_I = 0$) that factor investing has no advantage over direct investing. This is also intuitive that a large amount of informed trading in market factor brings up the information embedded into the price and squeezes the profiting space for factor funds.

The coexistence of (i) informed and uninformed households and (ii) direct investing and fundamental investing implies that

$$\gamma\,(u_a - u_b) = \log\Big(\frac{e^{\gamma\,c} - \mu}{1-\mu}\Big) = \gamma\,(k + c).$$

Thus μ is determined as in (4.3.13). Combining the above results and recalling the relative

utility relation, I determine λ_a through

$$\frac{1}{2}\log\left(1+\frac{\gamma^2\,\beta_{v,e}\,\sigma_\varepsilon^2}{\lambda_a^2}\,\sigma_{q_2}^2\right)=\gamma\,(k+c),$$

yielding the equilibrium capital allocated to active funds given in the proposition.

Proof of Lemma 4.3.3 The feasibility of the fundamental-dominating equilibrium within the region specified by condition (4.3.21) is not hard to justify. In the proof I mainly show the continuity at the switching point.

Upon reaching the peak level of factor investing, the delegation dominating equilibrium implies that

$$\lambda_b=\frac{\gamma\,\sigma_{q_2}}{\mu}\sqrt{\frac{1}{\left(\frac{1}{\sigma_e^2}+\frac{1}{\sigma_\varepsilon^2}\right)\left(e^{2\,\gamma\,(k+c)}-1\right)}}=\frac{1}{\mu+1},$$

and

$$\lambda_a=1-\lambda_b=\frac{\mu}{1+\mu}.$$

Based on the results in Proposition 4.3.2, we shall have

$$\alpha_b=\frac{k\big[1-\,(1+\mu)\,\lambda_b\big]}{2\,\lambda_b-1}=0,\qquad \alpha_a=\frac{k\,(1-\mu)\,\lambda_b}{2\,\lambda_b-1}=\frac{k\,(1-\mu)}{1-\mu}=k;$$

and

$$\xi=\frac{\big[2\,\lambda_b-1\big]\,\mu}{\big[1-\lambda_b\big]\,(1-\mu)}=\frac{\big[2-(\mu+1)\big]\,\mu}{(\mu+1-1)\,(1-\mu)}=1.$$

The equilibrium amount of skilled active managers, μ, remains unchanged as in (4.3.13) and is determined by the information frictions and risk appetites. Condition (4.3.15) is degenerate in this case and is automatically satisfied. At this upper bound, factor investing has reached its maximal market share, with the marginal value added by the factor investors before fees approaching zero, mathematically, $u_b-u_h\to 0$.

This, with factor investing driven to non-profitable boundary, is connected to the fundamental-dominating equilibrium, in which case alongside this boundary line we have

$$\lambda_a=\gamma\,\sigma_{q_2}\sqrt{\frac{1}{\left(\frac{1}{\sigma_e^2}+\frac{1}{\sigma_\varepsilon^2}\right)\left(e^{2\,\gamma\,(k+c)}-1\right)}}=\frac{\mu}{\mu+1},$$

where the first equality comes from the results in Proposition 4.3.3 and the second equality comes from the boundary condition implied by (4.3.20) and (4.3.21). Therefore, I show that alongside the boundary, any choice of (λ_b,λ_h) satisfying $\lambda_b+\lambda_h=1/(1+\mu)$ forms an equilibrium. This is because households treat direct investing and factor investing equally.

I can now conclude that the "switching" between equilibria at the boundary is always upper-semicontinuous.

Proof of Proposition 4.5.1 From equation (D.3.15), the demand for factor investing in the interior equilibrium is

$$\lambda_b = \frac{\gamma\,\sigma_\varepsilon\,\sigma_{q_2}}{\mu}\sqrt{\frac{\beta_{v,e}}{e^{2\,\gamma\,(k+c)}-1}},$$

where μ is a function of information costs and risk appetite, given by expression (4.3.13).

(i) Taking derivative of λ_b with respect to household searching cost c yields

$$\frac{\partial\lambda_b}{\partial c} = -\frac{\gamma\,e^{-\gamma\,c}\,(e^{2\,\gamma\,(k+c)}-e^{\gamma\,(k+c)}-1)}{(e^{\gamma\,k}-e^{-\gamma\,c})\,(1+e^{\gamma\,(k+c)})}\,\lambda_b,$$

which is positive if $c < \frac{1}{\gamma}\log\left(\frac{(1+\sqrt{5})\,e^{-\gamma\,k}}{2}\right)$ and negative otherwise. Thus λ_b is of a humped shape with respect to searching cost c.

(ii) Taking derivative of λ_b with respect to managers' information cost k yields

$$\frac{\partial\lambda_b}{\partial k} = -\frac{\gamma\,e^{-\gamma\,(k-c)}\,(e^{2\,\gamma\,(k+c)}-e^{\gamma\,(k+c)}+e^{\gamma\,c}-1)}{(e^{\gamma\,k}-1)\,(e^{\gamma\,k}-e^{-\gamma\,c})\,(1+e^{\gamma\,(k+c)})}\,\lambda_b < 0.$$

Hence, λ_b is monotonically decreasing in k.

(iii) Taking derivative of λ_b with respect to level of risk aversion γ yields

$$\begin{aligned}\frac{\partial\lambda_b}{\partial\gamma} = -\frac{e^{-\gamma\,c}\,\sigma_{q_2}\,\sigma_\varepsilon}{\sqrt{2}\,(e^{\gamma\,k}-1)^2\,(1+e^{\gamma\,(k+c)})}\sqrt{\beta_{v,e}\left(\frac{e^{\gamma\,(k+c)}+e^{-\gamma\,(k+c)}}{e^{\gamma\,(k+c)}-e^{-\gamma\,(k+c)}}\right)}\Big(1-\gamma\,c-e^{\gamma\,(k+c)}\,\gamma\,(k+c)\\ +e^{\gamma\,(2\,k+c)}\,\gamma\,(k+c)+e^{2\,\gamma\,(k+c)}\,(c\,\gamma-1)+(e^{\gamma\,(2\,c+3\,k)}-e^{\gamma\,k})\,(1-c\,\gamma-k\,\gamma)\Big).\end{aligned}$$

Hence, how λ_b varies with respect to γ depends on the zeros of expression

$$1-\gamma\,c-e^{\gamma\,(k+c)}\,\gamma\,(k+c)+e^{\gamma\,(2\,k+c)}\,\gamma\,(k+c)+e^{2\,\gamma\,(k+c)}\,(c\,\gamma-1)+(e^{\gamma\,(2\,c+3\,k)}-e^{\gamma\,k})\,(1-c\,\gamma-k\,\gamma).$$

(iv) Taking derivative of λ_b with respect to individual supply uncertainty $\sigma^2_{q_2}$ yields

$$\frac{\partial\lambda_b}{\partial\sigma^2_{q_2}} = \frac{e^{-\gamma\,k}-e^{-\gamma\,c}}{(e^{\gamma\,k}-1)}\,\gamma\,\sigma_\varepsilon\sqrt{\frac{\beta_{v,e}}{e^{2\,\gamma\,(k+c)}-1}} > 0.$$

Similarly by taking derivative of λ_b with respect to idiosyncratic risk σ^2_e, one can see that λ_b is monotonically increasing in $\sigma^2_{q_2}$ and σ^2_e.

(v) For this part I only have numerical results, as discussed in Appendix Section D.4.

Proof of Lemma 4.5.1

(i) Taking λ_h as given, I mainly investigate the influence on price terms brought by the change of $r_{f-f} = \lambda_b/\lambda_a$. Because the total demand is normalized to be a unit, we have $\lambda_a = (1-\lambda_h)/(1+r_{f-f})$. Thus ϕ_q can be rewritten as

$$\phi_q = \frac{\gamma\,\sigma_\varepsilon^2\,(1+r_{f-f})}{\big(\xi+\mu\,(1-\xi)\big)\,(1-\lambda_h)}. \tag{D.3.17}$$

If the effect on r_{f-f} dominates the change in direct investing, λ_h, then a change of demand for factor investing affects ϕ_q directly.

(a) Following the results in Proposition 4.2.1, the expected price of market index, after rearranging, is

$$\mathbb{E}[p_I] = \theta_0 = \overline{s} - \frac{\overline{q}_I\,\gamma\,Var(S|s)}{(\lambda_b+\lambda_a)\left(1-\frac{Var(S|s)}{Var(S|p_I)}\right)+\frac{Var(S|s)}{Var(S|p_I)}}.$$

Note that $\frac{Var(S|s)}{Var(S|p_I)} \in (0,1)$. $Var(S|p_I)$ is a function of θ_q. Thus the expected price is only affected by the total amount of capital in funds. So does the expected price of individual asset if it is positively exposed to the aggregate market risk, because

$$\mathbb{E}[p_i] = \beta_i\,\mathbb{E}[p_I], \qquad \text{for } i = 1,2,...,l.$$

The variance of index price is,

$$Var(p_I) = \theta_b^2\,\sigma_s^2 + \theta_b^2\,\theta_q^2\,\sigma_{q_I}^2,$$

which also remains unchanged provided that $\lambda_b+\lambda_a$ is unaffected. The variance of stock-picking asset price, for instance the first stock-picking asset with payoff $\mathcal{E}_2$, is

$$Var(\tilde{p}_2) = \phi_a^2\,\sigma_e^2 + \phi_a^2\,\phi_q^2\,\sigma_{q_2}^2.$$

If a rise in λ_b mainly increases r_{f-f} (decreases λ_a), then ϕ_q is larger according to equation (D.3.17), with the rest coefficients remaining unaffected. Hence $Var(\tilde{p}_2)$ increases. The effect of r_{f-f} on variance is approximately quadratic.

I call a market with more indexation a more "concentrated" market. Note that when an idiosyncratic shock occurs ($\sigma_{q_2}^2$, measuring the size of the shock, varies), a more concentrated industry is more exposed, due to the increased sensitivity towards the shock implied by the variance analysis.

(b) I calculate the covariance of asset-specific components using the definition, i.e.,

$$Cov(\tilde{p}_2,\tilde{p}_3) = \phi_a^2\,Cov(e_2,e_3) + \rho_{23}\,\phi_a^2\,\phi_q^2\,\sigma_{q_2}^2.$$

Because we have $Cov(e_2, e_3) = Cov(\mathcal{E}_2 + \varepsilon_2, \mathcal{E}_3 + \varepsilon_3) = Cov(\mathcal{E}_2, \mathcal{E}_3) = \rho_{23}\,\sigma_e^2$, the covariance is of the same sign as ρ_{23}. The covariance moves in a similar fashion as the variance of individual asset's price. The expansion of factor delegation, in other words, the shrink of active delegation, drives up ϕ_q, thus the covariance between the prices of asset-idiosyncratic loadings rises (in the absolute value sense).[2] The effect of r_{f-f} on covariance is approximately quadratic.

(c) The conclusion comes directly from result (D.3.10) and (D.3.11) in Lemma D.3.1.

(ii) Internalizing λ_h, we have $\lambda_h = r_{n-f}/(1 + r_{n-f})$ and $\lambda_b + \lambda_a = 1/(1 + r_{n-f})$.

(a) Following the results in Proposition 4.2.1, we know that $\mathbb{E}[p_I] = \theta_0$. Differentiating it with respect to r_{n-f} yields

$$\frac{\partial \theta_0}{\partial r_{n-f}} = -\frac{\overline{q}_I\,\gamma^3\,\sigma_{q_I}^2\,\sigma_s^4\,\sigma_\varepsilon^4\left(1 + 2\,r_{n-f} + \gamma^2\,(1 + r_{n-f})^2\,\sigma_{q_I}^2\,\sigma_\varepsilon^2\right)}{\left(\sigma_\varepsilon^2 + \gamma^2\,(1 + r_{n-f})^2\,\sigma_{q_I}^2\,\sigma_\varepsilon^4 + \sigma_s^2\,(1 + \gamma^2\,(1 + r_{n-f})\,\sigma_{q_I}^2\,\sigma_\varepsilon^2)\right)^2} < 0.$$

Hence, a drop in the ratio r_{n-f} increases the expected prices for both the market portfolio and individual assets.

(b) I rewrite the coefficients with respect to the ratio r_{n-f}, as follows,

$$\begin{aligned}
\theta_q &= \gamma\,(1 + r_{n-f})\,\sigma_\varepsilon^2,\\
\theta_b &= \frac{\sigma_s^2\left(1 + \gamma^2\,(1 + r_{n-f})\,\sigma_{q_I}^2\,\sigma_\varepsilon^2\right)}{\sigma_\varepsilon^2 + \gamma^2\,(1 + r_{n-f})^2\,\sigma_{q_I}^2\,\sigma_\varepsilon^4 + \sigma_s^2\left(1 + \gamma^2\,(1 + r_{n-f})\,\sigma_{q_I}^2\,\sigma_\varepsilon^2\right)}.
\end{aligned}$$

Notice that

$$\frac{\partial \theta_0}{\partial r_{n-f}} = \overline{q}_I\,\gamma\,\sigma_s^2\,\frac{\partial \theta_b}{\partial r_{n-f}}.$$

As r_{n-f} decreases, θ_q decreases and θ_b increases. Therefore, the effect on variance $Var(p_I) = \theta_b^2\,\sigma_s^2 + \theta_b^2\,\theta_q^2\,\sigma_{q_I}^2$, at a first glance, turns out to be mixed. In fact, differentiating the variance with respect to r_{n-f} yields

$$\begin{aligned}
&2\,\gamma^2\,\sigma_{q_I}^2\,\sigma_s^4\,\sigma_\varepsilon^4\big[-r_{n-f}\,\sigma_s^2 + (1 + r_{n-f})\,(1 + \gamma^2\,\sigma_{q_I}^2\,\sigma_s^2)\,\sigma_\varepsilon^2 +\\
&\quad\gamma^2\,(1 + r_{n-f})^2\,\sigma_{q_I}^2\,(2 - r_{n-f} + 2\,\gamma^2\,(1 + r_{n-f})\,\sigma_{q_I}^2\,\sigma_s^2)\,\sigma_\varepsilon^4 +\\
&\quad\gamma^4\,(1 + r_{n-f})^3\,\sigma_{q_I}^4\,(1 - r_{n-f} + \gamma^2\,(1 + r_{n-f})\,\sigma_{q_I}^2\,\sigma_s^2)\,\sigma_\varepsilon^6\big]\\
&\quad/\left(\sigma_\varepsilon^2 + \gamma^2\,(1 + r_{n-f})^2\,\sigma_{q_I}^2\,\sigma_\varepsilon^4 + \sigma_s^2\,(1 + \gamma^2\,(1 + r_{n-f})\,\sigma_{q_I}^2\,\sigma_\varepsilon^2)\right)^3.
\end{aligned}$$

The denominator is always positive. When the information is relatively precise, i.e., σ_ε^2 is tiny, the first term in the bracket of numerator dominates, as the rest

[2] Glosten et al. (2016) point out that the increasing co-movement can be partly driven by fundamental factors, which lines up with this situation in my model. Whereas other traditional literature show a non-fundamental factor-driven results (e.g., Harris and Gurel, 1986; Vijh, 1994; Barberis et al., 2005; Peng and Xiong, 2006; Greenwood, 2007; Staer, 2014; Da and Shive, 2018).

are the higher orders of σ_ε^2. Thus in this case the variance of index price is negatively correlated with the ratio r_{n-f}.

(c) I calculate the covariance using the definition, i.e.,

$$Cov(p_2, p_3) = Cov(\tilde{p}_2 + \beta_2\, p_I, \tilde{p}_3 + \beta_3\, p_I) = \phi_a^2\, Cov(e_2, e_3) + \rho_{23}\, \phi_a^2\, \phi_q^2\, \sigma_{q_2}^2 + \beta_2\, \beta_3\, Var(p_I).$$

If the index is traded actively enough, i.e., if $\sigma_{q_I}^2$ dominates, then $Cov(p_2, p_3)$ moves in the same way as $Var(p_I)$.

Proof of Lemma 4.5.2 Simply taking derivatives of η_I and $\tilde{\eta}_2$ over r_{f-f} yields the results. From the relation (D.3.7), one can see that η_2 moves in the same direction as $\tilde{\eta}_2$. Note that the effect of r_{f-f} on price efficiency is more sophisticated, rather than quadratic.

D.4 Heterogeneous Households

D.4.1 Who favors factor investing?

Each household, with statistical structure $\{W^n, c^n, \gamma^n\}$, will employ the strategy

$$\tilde{\boldsymbol{x}}_h^n = \left(\frac{\mathbb{E}[S|p_I] - p_I}{\gamma^n\, Var(S|p_I)},\, 0 \right)^\intercal,$$

if she chooses to invest herself. The funds' strategies are independent of n and identical to all the households. However, because wealth endowment influences risk appetite, the strategies generate heterogeneous certainty-equivalent utility to heterogeneous households. Following the analysis in Lemma D.3.1, we have the following certainty-equivalent payoffs to household n from different investment vehicles, denoted by u_j^n $(j = a, p, h)$,

$$\begin{aligned}
u_a^n &= \frac{1}{2\,\gamma^n} \left[\log\Big(\frac{Var(S - p_I)}{Var(S|s)}\Big) + \log\Big(\frac{Var(\mathcal{E}_2 - \tilde{p}_2)}{Var(\mathcal{E}_2|e_2)}\Big) + \frac{(\overline{s} - \theta_0)^2}{Var(S - p_I)} \right], \\
u_b^n &= \frac{1}{2\,\gamma^n} \left[\log\Big(\frac{Var(S - p_I)}{Var(S|s)}\Big) + \frac{(\overline{s} - \theta_0)^2}{Var(S - p_I)} \right], \\
u_h^n &= \frac{1}{2\,\gamma^n} \left[\log\Big(\frac{Var(S - p_I)}{Var(S|p_I)}\Big) + \frac{(\overline{s} - \theta_0)^2}{Var(S - p_I)} \right].
\end{aligned}$$

Notice that $\gamma^n\, u_j^n$ is independent of n.

A household weakly prefers skilled active manager to investing herself if and only if

$$u_a^n - \alpha_a^n - c^n \geq u_h^n, \quad \Rightarrow \quad \gamma^n\,(u_a^n - \alpha_a^n - c^n) \geq \gamma^n\, u_h^n. \tag{D.4.18}$$

Because the fees charged by funds are proportional to a household's wealth, a household is indifferent between direct investing and factor investing. Thus one can figure out the equilibrium fees of factor funds, as follows,

$$u_b^n - \alpha_b^n = u_h^n, \quad \Rightarrow \quad \gamma^n \, \alpha_b^n = \gamma^n \, (u_b^n - u_h^n) = \eta_I, \quad \Rightarrow \quad \alpha_b^n = \frac{\eta_I}{\gamma^R} \, W^n.$$

Similar to relation (4.3.11), we have $(\alpha_a^n - \alpha_b^n)/W^n = k$, where k is the manager's information cost per unit of wealth. Then the equilibrium active delegation fees for household n become

$$\alpha_a^n = \Big(\frac{\eta_I}{\gamma^R} + k \Big) \, W^n.$$

Similarly, the model implies that households cannot gain from selecting active managers based on prior information, i.e.,

$$\mu \, e^{-\gamma^n \, (u_a^n - \alpha_a^n)} + (1 - \mu) \, e^{-\gamma^n \, (u_b^n - \alpha_a^n)} = e^{-\gamma^n \, u_h^n},$$

which, performing necessary calculation, yields

$$e^{-\gamma^n \, (u_a^n - u_b^n - k \, W^n)} = \frac{1 - e^{\gamma^R \, k} \, (1 - \mu)}{\mu}.$$

Linking the result to condition (D.4.18), (D.4.18) can be reduced to a feasible region determined by μ:

$$\gamma^n \, (u_a^n - u_b^n - k \, W^n) = - \log \Big(\frac{1 - e^{\gamma^R \, k} \, (1 - \mu)}{\mu} \Big) \geq \gamma^R \, \frac{c^n}{W^n}. \tag{D.4.19}$$

Thus, households with either larger wealth or higher cognitive ability are more likely to choose fundamental investing. While the rest, termed unsophisticated households, stay uninformed and favor more conservative investment vehicles. Up till now I have proved condition (4.4.23) in Lemma 4.4.1.

D.4.2 Equilibrium price and price efficiency

Because the wealth distribution across the population of households is unequal, each individual could contribute differently to the funds inflow due to their wealth effect. Therefore, I first figure out the equilibrium effect on market clearing, asset price, and price efficiency. Once these terms are determined, they shall be shared by all the households. In other words, defining $\mathbb{E}[\frac{1}{\gamma^n}] = \frac{1}{\gamma}$, the equilibrium price and price efficiency are independent of n.

The market clearing condition with wealth effect becomes

$$\frac{1}{W}\Big(\lambda_h\,\mathbb{E}[W^n\,\tilde{\boldsymbol{x}}_h^n|\text{uninformed}]+\big(\lambda_b+(1-\mu)\,(1-\xi)\,\lambda_a\big)\,\mathbb{E}[W^n|\text{uninformed}]\,\tilde{\boldsymbol{x}}_b \\ +\big(\xi\,\mathbb{E}[W^n|\text{informed}]+\mu\,(1-\xi)\,\mathbb{E}[W^n|\text{uninformed}]\big)\,\lambda_a\,\tilde{\boldsymbol{x}}_a\Big)=\tilde{\boldsymbol{q}},$$

scaling for expected wealth. Correspondingly, the coefficients of equilibrium price become

$$\theta_q=\frac{\gamma^R\,\sigma_\varepsilon^2}{\lambda_b\,\mathbb{E}[W^n|\text{uninformed}]+\lambda_a\,W},$$
$$\phi_q=\frac{\gamma^R\,\sigma_\varepsilon^2}{\big(\xi\,\mathbb{E}[W^n|\text{informed}]+\mu\,(1-\xi)\,\mathbb{E}[W^n|\text{uninformed}]\big)\,\lambda_a}.$$

Recall the definition $\beta_{v,s}=\frac{\sigma_s^2}{\sigma_s^2+\sigma_\varepsilon^2}$ and $\beta_{v,e}=\frac{\sigma_e^2}{\sigma_e^2+\sigma_\varepsilon^2}$. Hence, the price inefficiency of market portfolio, η_I, becomes

$$\eta_I=-\frac{1}{2}\,\log\left(1-\frac{\beta_{v,s}\,\sigma_\varepsilon^2\,\sigma_{q_I}^2}{\big((\lambda_b\,\mathbb{E}[W^n|\text{uninformed}]+\lambda_a\,W)/\gamma^R\big)^2+\sigma_\varepsilon^2\,\sigma_{q_I}^2}\right). \tag{D.4.20}$$

Following the same rationale in Lemma D.3.1, one can characterize the relative utility in the heterogeneous world. For factor funds, we have

$$\gamma^n\,(u_b^n-u_h^n)=\eta_I.$$

For fundamental investors, we have

$$\gamma^n\,(u_a^n-u_h^n)=\eta_I+\frac{1}{2}\,\log\left(1+\frac{(\gamma^R)^2\,\beta_{v,e}\,\sigma_\varepsilon^2\,\sigma_{q_2}^2}{\big(\xi\,\mathbb{E}[W^n|\text{informed}]+\mu\,(1-\xi)\,\mathbb{E}[W^n|\text{uninformed}]\big)^2\,\lambda_a^2}\right). \tag{D.4.21}$$

D.4.3 Asset market equilibrium

The equilibrium amount of informed households, ξ, is determined endogenously given condition (D.4.19):

$$\xi=\mathbb{P}(u_a^n-c^n-\alpha_a^n\geq u_h^n)=\mathbb{P}\bigg(\gamma^n\,(u_a^n-u_b^n-k\,W^n)\geq\gamma^R\,\frac{c^n}{W^n}\bigg)$$
$$=\mathbb{P}\bigg(c^n\leq-\frac{1}{\gamma^R}\,\log\Big(\frac{1-e^{\gamma^R k}\,(1-\mu)}{\mu}\Big)\,W^n\bigg).$$

Given the distribution of $\{W^n,c^n\}$, with a joint density function $f_{W^n,c^n}(\cdot,\cdot)$ (separable if the two variables are independent), the probability ξ is related to the region determined

by μ, which is another endogenous value. Denote the region created by condition (D.4.19) and the support of distribution $f_{W^n,c^n}(\cdot,\cdot)$ as $\Omega(\mu)$. Then mathematically,

$$\xi = \mathbb{P}\big((W^n, c^n) \in \Omega(\mu) \big) = \iint_{\Omega(\mu)} f_{W^n,c^n}(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

which implies that ξ is a function of μ.

Now let us turn to the fund side. First, the skilled managers attract all the inflows from informed households and part of the inflows from the uninformed ones. Whereas the closet indexers can only gain by fooling the uninformed households. Thus the coexistence of skilled and opportunistic managers implies

$$\begin{aligned}\left(\mathbb{E}[\alpha_a^n - k\,W^n|\text{informed}]\,\frac{\xi}{\mu} + \mathbb{E}[\alpha_a^n - k\,W^n|\text{uninformed}]\,\frac{\mu\,(1-\xi)}{\mu} \right)\lambda_a \\ = \mathbb{E}[\alpha_a^n\,W^n|\text{uninformed}]\,\frac{(1-\mu)\,(1-\xi)}{1-\mu}\,\lambda_a,\end{aligned}$$

which, after simplification, yields

$$\frac{\eta_I}{\gamma^R\,\mu}\,\mathbb{E}[W^n|\text{informed}]\,\xi = k\,\mathbb{E}[W^n|\text{uninformed}]\,(1-\xi). \tag{D.4.22}$$

Note that

$$\mathbb{E}[W^n|\text{informed}]\,\xi = \mathbb{E}\big[W^n|(W^n,c^n)\in\Omega(\mu)\big]\,\mathbb{P}\big((W^n,c^n)\in\Omega(\mu) \big) = \iint_{\Omega(\mu)} x\,f_{W^n,c^n}(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

and similar for $\mathbb{E}[W^n|\text{uninformed}]\,(1-\xi)$. Then through finding the fixed point of equation (D.4.22) I can express μ as a function of η_I. The household investment in finding out managers, ξ, determined by μ, becomes a function of η_I as well.

Second, the coexistence of factor and active funds shows that

$$\mathbb{E}[\alpha_b^n\,W^n|\text{uninformed}]\,\lambda_b = \mathbb{E}[\alpha_a^n\,W^n|\text{uninformed}]\,\frac{(1-\mu)\,(1-\xi)}{1-\mu}\,\lambda_a,$$

which implies

$$\frac{\lambda_b}{\lambda_a} = \frac{\eta_I + \gamma^R\,k}{\eta_I}\,(1-\xi).$$

Recall the form of η_I given by equation (D.4.20). With the above ratio of capital in factor

over fundamental investing, one can express λ_a using η_I, as follows,

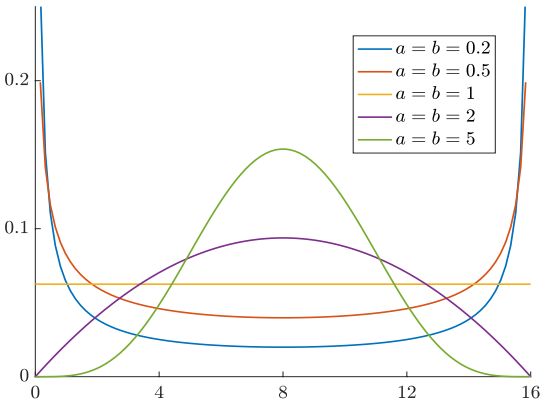
$$W\,\lambda_a + \mathbb{E}[W^n|\text{uninformed}]\,\lambda_b = \Big(\frac{\eta_I + \gamma^R\,k}{\eta_I}\,(1-\xi)\,\mathbb{E}[W^n|\text{uninformed}] + W\Big)\,\lambda_a$$
$$= \gamma^R\,\sigma_\varepsilon\,\sigma_{q_I}\sqrt{\frac{\beta_{v,s}}{1-e^{-2\,\eta_I}} - 1}.$$

Finally, I rely on the result of relative utility, equation (D.4.21), to solve the fixed point η_I. Similar to equation (D.3.14) in the proof of Proposition 4.3.1, and combining condition (D.4.19), we have

$$\gamma^n\,(u_a^n - u_b^n) = \frac{1}{2}\,\log\left(1 + \frac{(\gamma^R)^2\,\beta_{v,e}\,\sigma_\varepsilon^2\,\sigma_{q_2}^2}{\big(\xi\,\mathbb{E}[W^n|\text{informed}] + \mu\,(1-\xi)\,\mathbb{E}[W^n|\text{uninformed}]\big)^2\,\lambda_a^2}\right)$$
$$= \gamma^R\,k - \log\Big(\frac{1-e^{\gamma^R\,k}\,(1-\mu)}{\mu}\Big),$$

where μ, ξ, and λ_a are all functions of η_I. Therefore, one can determine the equilibrium price informativeness, then the rest terms follow.

Lemma 4.4.2 Before showing the result, I first give the details of how to measure wealth dispersion. As shown in the main body, I vary parameter a of beta distribution, Beta(a, a), holding the mean and support constant. A higher value of a implies a more concentrated statistical distribution, intuitively, shown in Figure D.4. Taking Gini coefficient as a measure of wealth dispersion, a higher a maps into a lower Gini coefficient.



a = b	Gini
0.2	0.47
0.5	0.41
1	0.33
2	0.26
5	0.17

Figure D.4: *Density of Beta distribution, Beta(a, b), for various parameters $a = b$, and the corresponding Gini index.*

Fixing the information cost c, the condition to determine ξ reduces to

$$\xi = \mathbb{P}\left(W^n \geq -\frac{\gamma^R c}{\log\left(\frac{1-e^{\gamma^R k}(1-\mu)}{\mu}\right)} \right).$$

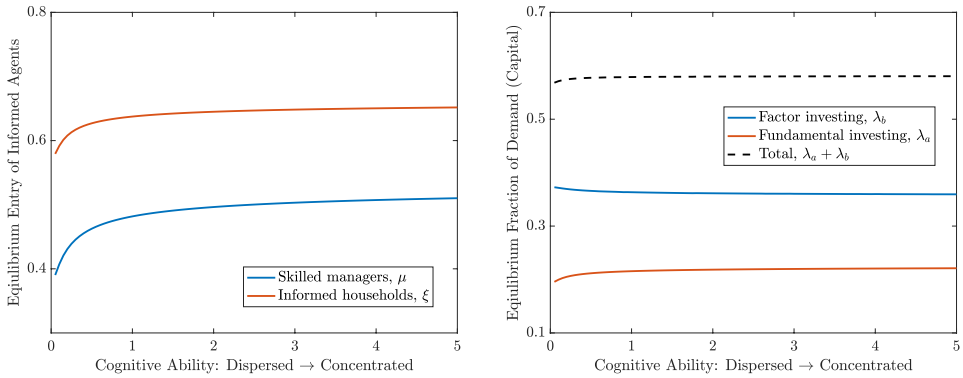
The rest of the analysis hold. Thus numerically, as I vary the choice of parameter a of Beta distribution to control wealth dispersion level, the asset market equilibrium terms vary correspondingly. An intuitive example and explanation are provided in the main body of the paper. It shows that higher wealth inequality engenders more skilled managers, less households willing to become informed, and higher capital in both funds. The outcomes are robust across different choices of exogenous variables.

Lemma 4.4.3 Lemma 4.4.3 follows exactly the same rationale. With a fixed wealth across the population of households $W^n \equiv W$, households investment in finding out manager quality, ξ, is determined by

$$\xi = \mathbb{P}\left(c^n \leq -\frac{W}{\gamma^R} \log\left(\frac{1-e^{\gamma^R k}(1-\mu)}{\mu}\right) \right) = \frac{B\left(-\frac{W}{\gamma^R}\log\left(\frac{1-e^{\gamma^R k}(1-\mu)}{\mu}; a, a\right)\right)}{B(a,a)},$$

where $B(\cdot\,; a, a)$ is the incomplete beta function and $B(a, a)$ is the beta function.

Cognitive ability, in this framework, does not affect certainty-equivalent wealth. Moreover, fixing the wealth level, households' contribution towards fund inflow is identical. Therefore I do not need to distinguish between demand and capital. A numerical example is given below. Through examining a series of reasonable choice of exogenous variables, one can tell that the effect is robust, yet of a tiny quantity.



A. Informed agents B. Demand/Capital

Figure D.5: *Cognitive ability inequality on capital allocation.* I vary the parameter of Beta distribution, a, in $(0, 5]$, holding the expected information cost $c = 0.3$ and its support $[0, 0.6]$. For the other parameters, take $\gamma^R = 3$, $k = 0.05$, $W = 5$, $\sigma^2_{q_I} = \sigma^2_{q_2} = \sigma^2_e = \sigma^2_s = 2\,\sigma^2_\varepsilon = 0.6$.

Lemma 4.4.4 The density of AMH copula, by calculation, for any given correlation parameter θ, is

$$f_\theta(x,y) = \frac{\partial \text{AMH}(\theta)}{\partial x\, \partial y} = \frac{4\,W\,c\left(2\,c\,(1-\theta)\,(2\,W\,(1-\theta)+x\,\theta)+y\,\theta\,(2\,W+x-2\,W\,\theta+x\,\theta)\right)}{\left(4\,W\,c-\theta\,(2\,W-x)\,(2\,c-y)\right)^3}.$$

Recall the boundary condition to classify sophisticated and unsophisticated households, equation (4.4.23),

$$c^n \leq -\frac{1}{\gamma^R}\,\log\Big(\frac{1-e^{\gamma^R\,k}\,(1-\mu)}{\mu}\Big)\,W^n.$$

The region created by this condition, Ω, is determined by the line crossing origin with slope $\delta := -\frac{1}{\gamma^R}\,\log(\frac{1-e^{\gamma^R\,k}\,(1-\mu)}{\mu})$, and the projected rectangular domain $(W^n, c^n) \in [0, 2\,W] \times [0, 2\,c]$. If the slope is weakly smaller than c/W, then the region is of a triangle shape, with ξ determined by

$$\xi = \iint_{\Omega(\mu)} f_\theta(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\,W}\int_0^{\delta\,x} f_\theta(x,y)\,\mathrm{d}y\,\mathrm{d}x,$$

and

$$\begin{aligned}
&\mathbb{E}[W^n|\text{informed}]\,\xi = \iint_{\Omega(\mu)} x\,f_\theta(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\,W}\int_0^{\delta\,x} x\,f_\theta(x,y)\,\mathrm{d}y\,\mathrm{d}x,\\
&\mathbb{E}[W^n|\text{uninformed}]\,(1-\xi) = W - \mathbb{E}[W^n|\text{informed}]\,\xi.
\end{aligned}$$

If the slope is larger than c/W, then the region is of a trapezoid shape, with ξ determined by

$$\xi = 1 - \int_0^{2\,c}\int_0^{\frac{y}{\delta}} f_\theta(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

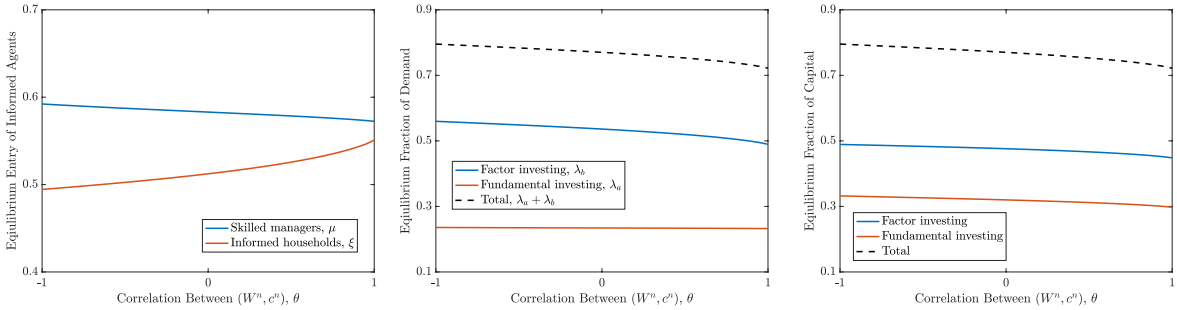
and

$$\begin{aligned}
&\mathbb{E}[W^n|\text{uninformed}]\,(1-\xi) = \int_0^{2\,c}\int_0^{\frac{y}{\delta}} x\,f_\theta(x,y)\,\mathrm{d}x\,\mathrm{d}y,\\
&\mathbb{E}[W^n|\text{informed}]\,\xi = W - \mathbb{E}[W^n|\text{uninformed}]\,(1-\xi).
\end{aligned}$$

Thus one can solve μ according to condition (D.4.22) as a function of η_I, and solve η_I correspondingly following the above analysis.

Numerically, the evidence that the correlation between wealth and cognitive ability influences the equilibrium capital allocation is given in Figure D.6. The outcomes are robust across different choices of exogenous variables: A negative θ implies that a wealthier household may have more insight towards a manager's quality. An unequal society stimulates demand for factor investing (Panel B and C of Figure D.6), as well as increases the compet-

itive pressure of active counterparts so that they have to lower their fees. The results are very much similar to the ones delivered by wealth inequality, which is reasonable because wealth effect dominates cognitive ability.



A. Informed agents B. Demand C. Capital

Figure D.6: *Unequal society on capital allocation.* I vary the parameter of AMH copula, θ, in $[-1, 1]$, holding the expected wealth $W = 5$ and information cost $c = 0.2$. For the other parameters, take $\gamma^R = 3$, $k = 0.05$, $\sigma_{q_I}^2 = \sigma_{q_2}^2 = \sigma_e^2 = \sigma_s^2 = 2\,\sigma_\varepsilon^2 = 0.5$.

D.4.4 Dynamics of opportunism managers

Proof of Example 4.4.1 I mainly derive the results for the heterogeneous equilibrium ("talent" market). I always use overscore for larger values and underscore for smaller values in notations. For the wealthy and informed households, with mass ξ, their properties and certainty-equivalent utilities can be characterized by the following set,

$$\underline{c},\ \overline{W},\ \underline{\gamma} = \frac{\gamma^R}{\overline{W}},\ \overline{u}_a,\ \overline{u}_b,\ \overline{u}_h,\ \overline{\alpha}_b = \frac{\eta_I}{\gamma^R}\,\overline{W},\ \overline{\alpha}_a = \left(\frac{\eta_I}{\gamma^R} + k\right)\overline{W}.$$

For the rest uninformed households, the corresponding set becomes

$$\overline{c},\ \underline{W},\ \overline{\gamma} = \frac{\gamma^R}{\underline{W}},\ \underline{u}_a,\ \underline{u}_b,\ \underline{u}_h,\ \underline{\alpha}_b = \frac{\eta_I}{\gamma^R}\,\underline{W},\ \underline{\alpha}_a = \left(\frac{\eta_I}{\gamma^R} + k\right)\underline{W}.$$

Note that we have $\overline{\gamma}\,\underline{u}_j = \underline{\gamma}\,\overline{u}_j$. The average absolute risk aversion level is determined by $\gamma = \frac{\gamma^R}{\mathbb{E}[W^n]} = \frac{\gamma^R}{\xi\,\overline{W} + (1-\xi)\,\underline{W}}$.

The rich and insightful group of households, as shown in the analysis in the main body, strictly prefers skilled fundamental investors, mathematically,

$$\overline{u}_a - \overline{\alpha}_a - \underline{c} > \overline{u}_b - \overline{\alpha}_b = \overline{u}_h.$$

Correspondingly, for the other group of households, we have

$$\begin{aligned} \underline{u}_a - \underline{\alpha}_a - \overline{c} &< \underline{u}_b - \underline{\alpha}_b = \underline{u}_h, \\ \mu\, e^{-\overline{\gamma}\,(\underline{u}_a - \underline{\alpha}_a)} + (1-\mu)\, e^{-\overline{\gamma}\,(\underline{u}_b - \underline{\alpha}_a)} &= e^{-\overline{\gamma}\,\underline{u}_h}. \end{aligned}$$

Managers self-select into their roles, as before. However, the contribution of the two groups of households to the fund inflows is different due to their wealth effect, yielding

$$\begin{aligned} \left((\overline{\alpha}_a - k\,\overline{W})\,\frac{\xi}{\mu} + (\underline{\alpha}_a - k\,\underline{W})\,\frac{\mu\,(1-\xi)}{\mu} \right) \lambda_a &= \underline{\alpha}_a\,\frac{(1-\mu)\,(1-\xi)}{1-\mu}\,\lambda_a. \\ \underline{\alpha}_b\,\lambda_b &= \underline{\alpha}_a\,(1-\xi)\,\lambda_a. \end{aligned}$$

Rearranging the definition of price efficiency yields

$$\lambda_a\,W + \lambda_b\,\underline{W} = \gamma^R\,\sigma_\varepsilon\,\sigma_{q_I}\,\sqrt{\frac{\beta_{v,s}}{1-e^{-2\,\eta_I}} - 1}.$$

Combine it with the funds coexisting conditions, I can express μ and λ_a both as functions of η_I:

$$\begin{aligned} \mu &= \frac{\eta_I}{k\,\gamma^R}\,\frac{\overline{W}\,\xi}{\underline{W}\,(1-\xi)}, \\ \lambda_a &= \frac{\gamma^R\,\sigma_\varepsilon\,\sigma_{q_I}\,\eta_I}{(\eta_I + k\,\gamma^R)\,(1-\xi)\,\underline{W} + \eta_I\,W}\,\sqrt{\frac{\beta_{v,s}}{1-e^{-2\,\eta_I}} - 1}. \end{aligned}$$

Finally, I rely on the result of relative utility, Lemma D.3.1, to solve the fixed point η_I:

$$\overline{\gamma}\,(\underline{u}_a - \underline{u}_b) = \frac{1}{2}\,\log\left(1 + \frac{(\gamma^R)^2\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi\,\overline{W} + \mu\,(1-\xi)\,\underline{W}\,)^2\,\lambda_a^2}\,\sigma_{q2}^2\right) = \gamma^R\,k - \log\left(\frac{1 - e^{\gamma^R\,k}\,(1-\mu)}{\mu}\right),$$

Then performing the same calculation as before yields the equilibrium with heterogeneous households.

For the homogeneous model, all the previous analysis applies. The only difference is due to the set up of non-unit wealth level, $W = 2$. Then condition (4.3.11) becomes

$$(\alpha_a - \alpha_b)/W = k,$$

i.e., fees are proportional to the total wealth.

D.5 Rich Asset Market

For tractability I analyze the asset market equilibrium assuming $l = 2$ in the main body of the paper. However the rationale is applicable to any asset market with more than two underlying assets. Specially, in the real financial market, the indices usually consist of more than 100 stocks, for instance, S&P 500, Russell 2000, FTSE 100, etc. Therefore, I extend my analysis to cover the case where asset market is "rich" enough, i.e., l is large. In this appendix I basically show that all the analysis in the main body holds for a world with large number of assets, and use the model to bridge manager skill and CAPM.

Assume index asset has a composition of $b_i = 1/l$ for any $i \in \{1, 2, ..., l\}$. As the index averages out idiosyncratic risks, one feasible covariance matrix, symmetric and positive-definite, features an identical correlation between any two idiosyncratic loadings:

$$\rho = \frac{Cov(\mathcal{E}_i, \mathcal{E}_j)}{\sigma_e^2} = -\frac{1}{l-1}, \qquad \forall\, i \neq j.$$

For an l large enough, this approximately generates independency among idiosyncratic risks. Thus the underlying assets are only correlated in terms of their factor loadings.

D.5.1 Equilibrium prices

Because a large l implies $\rho \to 0$, we have approximately the following equilibrium asset prices.

Corollary D.5.1 (**Equilibrium prices**) *The index price remains the same as in Proposition 4.2.1:*

$$p_I = \theta_0 + \theta_b \left[(s - \overline{s}) - \theta_q (q_I - \overline{q}_I) \right];$$

whereas the individual asset price satisfies, approximately, for any $i \in \{1, 2, ..., l\}$,

$$p_i = \beta_i\, p_I + \phi_a (\, e_i - \phi_q\, q_i \,).$$

where q_i is the supply of $\mathcal{E}_i$, and $(\theta_0, \theta_b, \theta_q)$ and (ϕ_a, ϕ_q) are parameters determined in the proof.

With the equilibrium prices, one can recover the price efficiency as before. Given the capital delegated to funds, the index price contains the same information as given in equation (4.2.6). Provided that $\sigma_{q_i}^2 \equiv \sigma_q^2$, the price informativeness of idiosyncratic attributes,

equation (D.3.6), can be generalized to

$$\tilde{\eta}_i \coloneqq -\frac{1}{2}\log\Big(\frac{Var(\mathcal{E}_i|e_i)}{Var(\mathcal{E}_i|\tilde{p}_i)}\Big) = -\frac{1}{2}\log\left(1-\frac{\sigma_\varepsilon^2\,\sigma_e^2\,\sigma_q^2/(\sigma_e^2+\sigma_\varepsilon^2)}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2/\gamma^2+\sigma_\varepsilon^2\,\sigma_q^2}\right)\in(0,\infty),\quad \forall\, i.$$

To make a reasonable assumption, let fundamental investors be able to produce idiosyncratic information of m stocks ($m \leq n$). In a market with large number of underlying assets, it is fair that factor investors have weak signal of, say, 500 stocks thus are able to time the factor, while fundamental investors have strong signal towards 20 attributes of the individual assets in the basket. A proper choice of m captures the rationale in a stylized fashion. Following the same rationale, I relate price efficiency to relative utilities, with the approximate results specified in Corollary D.5.2.

Corollary D.5.2 (**Relative utility**) *The excess performance of fund industry, represented using the relative utility, can be linked to the price efficiency in the following ways:*

$$\gamma\,(u_b-u_h)=\eta_I,$$
$$\gamma\,(u_a-u_b)=\frac{m-1}{2}\log\Big(1+\frac{\gamma^2\,\sigma_\varepsilon^2\,\sigma_e^2\,\sigma_q^2/(\sigma_e^2+\sigma_\varepsilon^2)}{(\,\xi+\mu\,(1-\xi)\,)^2\,\lambda_a^2}\Big).$$

D.5.2 Asset market equilibrium

Now I discuss how a rich asset market affects the equilibrium terms. Recall the delegation fees charged by active managers, equation (4.3.11). It is fair that the information cost of active managers increases with the number of assets they emphasize on, m, yet is bounded from above as the marginal cost of investigating the final asset is approximately zero. Thus I update equation (4.3.11) to be $\alpha_a - \alpha_b = k(m)$. Provided with the previous analysis, I assume $k(m)$ is Lipschitz continuous.

Then with the main ideas remaining unchanged, I find that a relatively large m can increase the amount of factor investing, in fact, fundamental investing as well. Mathematically we have

$$\lambda_b = \frac{\gamma\,\sigma_{q_2}}{\mu}\sqrt{\frac{1}{\big(\frac{1}{\sigma_e^2}+\frac{1}{\sigma_\varepsilon^2}\big)\big(e^{\frac{2\,\gamma\,(k(m)+c)}{m-1}}-1\big)}}.$$

When there are many underlying assets, another outcome is the increasing index price efficiency. This follows directly from the increase in money invested in funds—the more informed trading conducted by the fund managers, the more information the final price reflects. Furthermore, though the entry of skilled managers, μ, is unrelated to the number of assets, the amount of informed households, ξ, increases due to the higher price

efficiency. On the one hand, with more informed households delegating wealth to skilled active managers, each individual asset contributes less excess return to the selected portfolio. On the other hand, more idiosyncratic loadings provide extra space for fundamental investors to profit. Therefore the negative and positive effects partly cancel out with each other. Finally, because the coexistence of factor and active funds always implies that $\big(\xi + \mu\,(1-\xi)\big)\,\lambda_a = \mu\,\lambda_b$ in an interior equilibrium, one can conclude that the price efficiency of individual assets also enhances.

All the rest analysis of equilibria remains unchanged. There still exists an upper bound of the fraction of capital in factor investing, yet the condition upon reaching the peak level varies with the number of assets.

D.5.3 Skilled active funds and alpha

I now discuss the outperformance of skilled active managers. One common way to estimate a fund's excess return is through the Capital Asset Pricing Model (CAPM) and look into the "alpha," i.e., the intercept of the Security Market Line (SML), that fund generates. The only group of investors who can capture the abnormal return (funds' return minus the market return) is the group of fundamental investors, who hold inside information of idiosyncratic loadings. They are known for charging higher fees due to the alpha they generate.

Recalling the results in Section 4.2.3, skilled active managers' strategy is $\boldsymbol{x}_a = (x_{1a}, x_{2a}, ..., x_{la})^\intercal$, specified by equation (4.2.4). Then the weight put on asset i by the skilled manager is

$$\omega_{ia} = \frac{x_{ia}\,p_i}{\boldsymbol{x}_a^\intercal \boldsymbol{p}}, \qquad i = 1, 2, ..., l.$$

The gross returns of asset i, R_i, and market portfolio, R_m, are

$$R_i = \frac{v_i}{p_i}, \qquad R_m = \frac{S}{p_I}, \qquad i = 1, 2, ..., l.$$

Define the vector of risky asset weight $\boldsymbol{\omega}_a = (\omega_{1a}, \omega_{2a}, ..., \omega_{la})^\intercal$ and the vector of risky asset return $\boldsymbol{R} = (R_1, R_2, ..., R_l)^\intercal$. Then the skilled funds' return, denoted by R_a, can be expressed as $R_a = \boldsymbol{\omega}_a^\intercal \boldsymbol{R}$. Because the risk-free rate is normalized to be one, I use $\mathbf{1}$ as the l-dimensional unit vector. In the following lemma I develop the model-implied CAPM for the expected excess fund return.

Lemma D.5.1 (**Model-implied CAPM**) *If the supply of systematic risk dominates,*

then the expected excess return of skilled active funds satisfies

$$\mathbb{E}[R_a] - 1 = \mathbb{E}[\,\boldsymbol{\omega}_a^{\intercal}\,(\boldsymbol{R} - \mathbf{1})\,] = \alpha + \beta\,(\mathbb{E}[R_m] - 1).$$

The intercept satisfies $\alpha = \sum_{i=1}^{l} \frac{1}{\gamma}\big(\frac{Var(\tilde{v}_i - \tilde{p}_i)}{Var(\tilde{v}_i|\mathcal{I}_a)} - 1\big) - A_i$*, where* $\tilde{v}_i$ *is the payoff of the orthogonal asset* i *defined in Lemma D.2.1 and* $\tilde{p}_i$ *is the corresponding price. The slope satisfies* $\beta = \sum_{i=1}^{l} \hat{\omega}_{ia}\,\frac{Cov(R_i, R_m)}{Var(R_m)}$*.* A_i *and* $\hat{\omega}_{ia}$ *are positive numbers and will be specified in the proof.*

Proof: The weight placed on asset i is defined to be $\omega_{ia} = \frac{x_{ia}\,p_i}{\boldsymbol{x}_a^{\intercal}\boldsymbol{p}}$. Note that the element will simply be zero if fundamental investors do not hold this asset. Thus the expected excess return satisfies

$$\begin{aligned}\mathbb{E}[R_a - 1] = \mathbb{E}[\,\boldsymbol{\omega}_a^{\intercal}\,(\boldsymbol{R} - \mathbf{1})\,] &= \sum_{i=1}^{l} \mathbb{E}\Big[\,\frac{x_{ia}\,p_i}{\boldsymbol{x}_a^{\intercal}\boldsymbol{p}}\,\Big(\frac{v_i}{p_i} - 1\Big)\,\Big] = \frac{1}{\boldsymbol{x}_a^{\intercal}\boldsymbol{p}}\sum_{i=1}^{l}\mathbb{E}[\,x_{ia}\,(v_i - p_i)\,] \\ &= \frac{1}{\boldsymbol{x}_a^{\intercal}\boldsymbol{p}}\sum_{i=1}^{l}\Big(\,\mathbb{E}[x_{ia}]\,\mathbb{E}[v_i - p_i] + Cov(x_{ia}, v_i - p_i)\,\Big).\end{aligned} \tag{D.5.23}$$

First, I work out the second term on the RHS—the sum of covariances. In matrix notation, $\sum_{i=1}^{n} Cov(x_{ia}, v_i - p_i) = \mathrm{Tr}(\,Cov(\boldsymbol{x}_a, \boldsymbol{v} - \boldsymbol{p})\,)$, where 'Tr' represents the trace of a matrix. If, by orthogonalization, I relate the payoff vector in original and orthogonal asset space, $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$, through a nonsingular transforming matrix Γ, i.e., $\boldsymbol{v} = \Gamma\,\tilde{\boldsymbol{v}}$, then we have

$$\mathrm{Tr}(\,Cov(\boldsymbol{x}_a, \boldsymbol{v} - \boldsymbol{p})\,) = \mathrm{Tr}(\,Cov(\,(\Gamma^{-1})'\tilde{\boldsymbol{x}}_a, \Gamma\,(\tilde{\boldsymbol{v}} - \tilde{\boldsymbol{p}})\,)\,) = (\Gamma^{-1}\Gamma)\,\mathrm{Tr}(\,Cov(\tilde{\boldsymbol{x}}_a, \tilde{\boldsymbol{v}} - \tilde{\boldsymbol{p}})\,).$$

Thus the sum of covariances can be simplified to

$$\begin{aligned}\sum_{i=1}^{l} Cov(x_{ia}, v_i - p_i) = \mathrm{Tr}(\,Cov(\tilde{\boldsymbol{x}}_a, \tilde{\boldsymbol{v}} - \tilde{\boldsymbol{p}})\,) &= \sum_{i=1}^{m} Cov(\tilde{x}_{ia}, \tilde{v}_i - \tilde{p}_i) \\ &= \sum_{i=1}^{m}\frac{1}{\gamma}\,\frac{Var(\mathbb{E}[\tilde{v}_i - \tilde{p}_i|\mathcal{I}_a])}{Var(\tilde{v}_i|\mathcal{I}_a)} = \sum_{i=1}^{m}\frac{1}{\gamma}\,\Big(\frac{Var(\tilde{v}_i - \tilde{p}_i)}{Var(\tilde{v}_i|\mathcal{I}_a)} - 1\Big),\end{aligned}$$

where the result lies on the strategy $\tilde{x}_{ia} = \frac{\mathbb{E}[\tilde{v}_i - \tilde{p}_i|\mathcal{I}_a]}{\gamma\,Var(\tilde{v}_i|\mathcal{I}_a)}$ and the law of total variance (D.3.8).

Next, consider the first term, i.e., the product of expectations. We have

$$\begin{aligned}\mathbb{E}[x_{ia}]\,\mathbb{E}[v_i - p_i] &= \mathbb{E}[x_{ia}]\,\mathbb{E}[p_i\,(R_i - 1)] \\ &= \mathbb{E}[x_{ia}]\,\big(\,\mathbb{E}[p_i]\,\mathbb{E}[R_i - 1] + Cov(p_i, R_i)\,\big) \\ &= \frac{Cov(R_i, R_m)}{Var(R_m)}\,\mathbb{E}[x_{ia}]\,\mathbb{E}[p_i]\,(\,\mathbb{E}[R_m] - 1\,) + \mathbb{E}[x_{ia}]\,Cov(p_i, R_i),\end{aligned}$$

where the last line holds approximately if the supply of index asset is relatively large, from $\mathbb{E}[R_i - 1] = \frac{Cov(R_i,R_m)}{Var(R_m)}\,(\,\mathbb{E}[R_m] - 1\,)$. The assumption is reasonable in that people usually regard index investment as where the liquidity is. Define $\hat{\omega}_{ia} = \frac{\mathbb{E}[x_{ia}]\,\mathbb{E}[p_i]}{\boldsymbol{x}_a^{\intercal}\boldsymbol{p}}$ and $A_i = \frac{\mathbb{E}[x_{ia}]\,Cov(p_i,R_i)}{\boldsymbol{x}_a^{\intercal}\boldsymbol{p}}$. Note that A_i is negative in that the price is negatively correlated with return ($Cov(p_i, R_i) < 0$).

Plugging the above results into expression of expected excess return, equation (D.5.23), then α and β follow.

Based on the results in Proposition 4.2.1, one has

$$Var(S|s) = \beta_{v,s}\,\sigma_\varepsilon^2,$$

and

$$Var(S - p_I) = (1 - \theta_b)^2\,\sigma_s^2 + \theta_b^2\,\sigma_\varepsilon^2 + \theta_b^2\,\theta_q^2\,\sigma_{p_I}^2.$$

The other return volatility and conditional variance terms follow similarly, especially when l is large, approximately,

$$Var(\tilde{v}_i|\mathcal{I}_a) = Var(\mathcal{E}_i|\mathcal{I}_a) = \beta_{v,e}\,\sigma_\varepsilon^2,$$

and

$$Var(\tilde{v}_i - \tilde{p}_i) = Var(\mathcal{E}_i - \tilde{p}_i) = \beta_{v,e}\,\sigma_\varepsilon^2\left(1 + \frac{\gamma^2\,\beta_{v,e}\,\sigma_\varepsilon^2}{(\,\xi + \mu\,(1 - \xi)\,)^2\,\lambda_a^2}\,\sigma_q^2\right).$$

Therefore, in any equilibrium, α is endogenously determined by the capital allocation strategy and investment in information acquisition.

As an example, I take a look at the case when delegation dominates, i.e., $\lambda_h = 0$. This is essential as in a rich asset market, I have already shown that the fraction of households using delegation greatly increases. It is highly likely that direct investment can be completely driven out of the market. When delegation dominates, the following ratio could be simplified,

$$\frac{Var(S - p_I)}{Var(S|s)} = 1 + \gamma^2\,\beta_{v,s}\,\sigma_\varepsilon^2\,\sigma_{q_I}^2.$$

Then the abnormal return α becomes

$$\alpha = \gamma\,\sigma_{\varepsilon}^{2}\left(\beta_{v,s}\,\sigma_{q_I}^{2} + \frac{(m-1)\,\beta_{v,e}\,\sigma_{q}^{2}}{(\,\xi + \mu\,(1-\xi)\,)^{2}\,\lambda_{a}^{2}}\right) - \sum_{i=1}^{l} A_i.$$

□

As can be seen from the proof, when the systematic (σ_s^2) or idiosyncratic risk (σ_e^2) increases, α increases. The abnormal return is also larger when the supply of assets becomes more volatile.

Bibliography

Admati, Anat R and Paul Pfleiderer, "Robust financial contracting and the role of venture capitalists," *The Journal of Finance*, 1994, *49* (2), 371–402.

Agarwal, Vikas and Narayan Y Naik, "Risks and portfolio decisions involving hedge funds," *Review of Financial Studies*, 2004, *17* (1), 63–98.

_ , Naveen D Daniel, and Narayan Y Naik, "Role of managerial incentives and discretion in hedge fund performance," *The Journal of Finance*, 2009, *64* (5), 2221–2256.

Akerlof, George A and Janet L Yellen, "The fair wage-effort hypothesis and unemployment," *The Quarterly Journal of Economics*, 1990, *105* (2), 255–283.

Alexander, Gordon J, Gjergji Cici, and Scott Gibson, "Does motivation matter when assessing trade performance? An analysis of mutual funds," *Review of Financial Studies*, 2007, *20* (1), 125–150.

Ali, Mir M, NN Mikhail, and M Safiul Haq, "A class of bivariate distributions including the bivariate logistic," *Journal of Multivariate Analysis*, 1978, *8* (3), 405–412.

Allen, Franklin, Ana Babus, and Elena Carletti, "Financial crises: Theory and evidence," *Annual Review of Financial Economics*, 2009, *1* (1), 97–116.

Anadu, Kenechukwu, Mathias Kruttli, Patrick E McCabe, Emilio Osambela, and Chaehee Shin, "The shift from active to passive investing: Potential risks to financial stability?," *Available at SSRN*, 2018.

Aragon, George O and Vikram Nanda, "Tournament behavior in hedge funds: High-water marks, fund liquidation, and managerial stake," *Review of Financial Studies*, 2012, *25* (3), 937–974.

Ash, Robert B., *Real Analysis and Probability*, Academic Press, 1972.

Aumann, Robert J, Michael Maschler, and Richard E Stearns, *Repeated Games with Incomplete Information*, MIT Press, 1995.

Axelson, Ulf, Per Strömberg, and Michael S Weisbach, "Why are buyouts levered? The financial structure of private equity funds," *The Journal of Finance*, 2009, *64* (4), 1549–1582.

Baquero, Guillermo and Marno Verbeek, "The convexity and concavity of the flow-performance relationship for hedge funds," *Available at SSRN 2336118*, 2013.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, "Comovement," *Journal of Financial Economics*, 2005, *75* (2), 283–317.

Basak, Suleyman and Anna Pavlova, "Asset prices and institutional investors," *American Economic Review*, 2013, *103* (5), 1728–58.

_ , _ , and Alexander Shapiro, "Optimal asset allocation and risk shifting in money management," *Review of Financial Studies*, 2007, *20* (5), 1583–1621.

Bates, Steve, "Forced ranking," *HR Magazine*, 2003, *48* (6), 62–62.

Baye, Michael R, Dan Kovenock, and Casper G De Vries, "The all-pay auction with complete information," *Economic Theory*, 1996, *8* (2), 291–305.

Bebchuk, Lucian A and Scott Hirst, "Index funds and the future of corporate governance: Theory, evidence, and policy," Technical Report, National Bureau of Economic Research 2019.

Bebchuk, Lucian Arye, Jesse M Fried, and David I Walker, "Executive compensation in America: Optimal contracting or extraction of rents?," Technical Report, National Bureau of Economic Research 2001.

Becker, Gary S., Kevin M. Murphy, and Iván Werning, "The equilibrium distribution of income and the market for status," *Journal of Political Economy*, 2005, *113* (2), 282–310.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, "Hedge fund stock trading in the financial crisis of 2007–2009," *Review of Financial Studies*, 2012, *25* (1), 1–54.

_ , _ , and _ , "Do ETFs increase volatility?," *The Journal of Finance*, 2018, *73* (6), 2471–2535.

Benjamin, Daniel J, Sebastian A Brown, and Jesse M Shapiro, "Who is 'behavioral'? Cognitive ability and anomalous preferences," *Journal of the European Economic Association*, 2013, *11* (6), 1231–1255.

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

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

Bhattacharya, Ayan and Maureen O'Hara, "Can ETFs increase market fragility? Effect of information linkages in ETF markets," *Effect of Information Linkages in ETF Markets*, 2018.

Biais, Bruno and Fany Declerck, "Liquidity, competition & price discovery in the European corporate bond market," 2007.

Birkhoff, Garret and Saunders MacLane, *A Survey of Modern Algebra*, 4th ed., Macmillan, 1977.

Bond, Philip and Diego Garcıa, "Informed trading, indexing, and welfare," Technical Report, University of Washington Working Paper 2017.

Brown, David C and Shaun William Davies, "Moral hazard in active asset management," *Journal of Financial Economics*, 2017, *125* (2), 311–325.

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

Buchanan, James M, Robert D Tollison, and Gordon Tullock, "Efficient rent seeking," *Toward a Theory of the Rent Seeking Society*, 1980, pp. 97–121.

Buraschi, Andrea, Robert Kosowski, and Worrawat Sritrakul, "Incentives and endogenous risk taking: A structural view on hedge fund alphas," *The Journal of Finance*, 2014, *69* (6), 2819–2870.

Busse, Jeffrey A, "Another look at mutual fund tournaments," *Journal of Financial and Quantitative Analysis*, 2001, *36* (01), 53–73.

_ , Amit Goyal, and Sunil Wahal, "Investing in a global world," *Review of Finance*, 2013, *18* (2), 561–590.

Busse, Jeffrey, Tarun Chordia, Lei Jiang, and Yuehua Tang, "Mutual fund trading costs and diseconomies of scale," *Working Paper, Research Collection Lee Kong Chian School of Business.*, 2017.

Capponi, Agostino, Paul Glasserman, and Marko Weber, "Swing pricing for mutual funds: Breaking the feedback loop between fire sales and fund redemptions," *Management Science*, 2020.

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

Carpenter, Jennifer N, "Does option compensation increase managerial risk appetite?," *The Journal of Finance*, 2000, *55* (5), 2311–2331.

Cecchetti, Stephen Giovanni and Kermit L Schoenholtz, "The world of ETFs," 2016.

Chen, Joseph, Eric N Hughson, and Neal Stoughton, "Strategic mutual fund tournaments," *Available at SSRN 2023805*, 2018.

_ , Harrison Hong, Ming Huang, and Jeffrey D Kubik, "Does fund size erode mutual fund performance? The role of liquidity and organization," *American Economic Review*, 2004, *94* (5), 1276–1302.

Chen, Qi, Itay Goldstein, and Wei Jiang, "Payoff complementarities and financial fragility: Evidence from mutual fund outflows," *Journal of Financial Economics*, 2010, *97* (2), 239–262.

Cheney, Ward and David Kincaid, *Linear Algebra: Theory and Applications*, Vol. 110, The Australian Mathematical Society, 2009.

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

_ and _ , "Do fire sales create externalities?," *Journal of Financial Economics*, 2019.

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

Chevalier, Judith and Glenn Ellison, "Career concerns of mutual fund managers," *The Quarterly Journal of Economics*, 1999, *114* (2), 389–432.

Chiappori, Pierre-André and Monica Paiella, "Relative risk aversion is constant: Evidence from panel data," *Journal of the European Economic Association*, 2011, *9* (6), 1021–1052.

Choi, Darwin, Bige Kahraman, and Abhiroop Mukherjee, "Learning about mutual fund managers," *The Journal of Finance*, 2016, *71* (6), 2809–2860.

Choi, Jaewon and Sean Seunghun Shin, "Liquidity sensitive trading and fire sales by corporate bond mutual funds," Technical Report, Working Paper, University of Illinois 2018.

Christoffersen, Susan Kerr, David K Musto, and Bilge Yilmaz, "High water marks in competitive capital markets," in "EFA 2009 Bergen Meetings Paper" 2013.

_ , Donald B Keim, and David K Musto, "Valuable information and costly liquidity: Evidence from individual mutual fund trades," *Available at SSRN 951501*, 2006.

Cipriani, Marco, Antoine Martin, Patrick E McCabe, and Bruno Maria Parigi, "Gates, fees, and preemptive runs," *FRB of New York Staff Report*, 2014, (670).

Coles, Jeffrey L, Davidson Heath, and Matthew Ringgenberg, "On index investing," *Available at SSRN 3055324*, 2018.

_ , Zhichuan Li, and Albert Y Wang, "Industry tournament incentives," *Review of Financial Studies*, 2018, *31* (4), 1418–1459.

Cong, Lin William and Douglas Xu, "Rise of factor investing: Asset prices, informational efficiency, and security design," in "29th Australasian Finance and Banking Conference" 2016.

Coval, Joshua and Erik Stafford, "Asset fire sales (and purchases) in equity markets," *Journal of Financial Economics*, 2007, *86* (2), 479–512.

Cremers, KJ Martijn and Antti Petajisto, "How active is your fund manager? A new measure that predicts performance," *Review of Financial Studies*, 2009, *22* (9), 3329–3365.

Cremers, Martijn, Miguel A Ferreira, Pedro Matos, and Laura Starks, "Indexing and active fund management: International evidence," *Journal of Financial Economics*, 2016, *120* (3), 539–560.

Cronqvist, Henrik and Rüdiger Fahlenbrach, "CEO contract design: How do strong principals do it?," *Journal of Financial Economics*, 2013, *108* (3), 659–674.

Cuñat, Vicente and Maria Guadalupe, "Globalization and the provision of incentives inside the firm: The effect of foreign competition," *Journal of Labor Economics*, 2009, *27* (2), 179–212.

Cuthbertson, Keith, Dirk Nitzsche, and Niall O'Sullivan, "UK mutual fund performance: Skill or luck?," *Journal of Empirical Finance*, 2008, *15* (4), 613–634.

Da, Zhi and Sophie Shive, "Exchange traded funds and asset return correlations," *European Financial Management*, 2018, *24* (1), 136–168.

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

DeMarzo, Peter M and Michael J Fishman, "Optimal long-term financial contracting with privately observed cash flows," *Available at SSRN 562204*, 2004.

_ **and Ron Kaniel**, "Relative pay for non-relative performance: Keeping up with the Joneses with optimal contracts," Working paper SSRN 2755856 2017.

_ **and Yuliy Sannikov**, "Optimal security design and dynamic capital structure in a continuous-time agency model," *The Journal of Finance*, 2006, *61* (6), 2681–2724.

Diamond, Douglas W and Philip H Dybvig, "Bank runs, deposit insurance, and liquidity," *Journal of Political Economy*, 1983, *91* (3), 401–419.

Dick-Nielsen, Jens, Peter Feldhütter, and David Lando, "Corporate bond liquidity before and after the onset of the subprime crisis," *Journal of Financial Economics*, 2012, *103* (3), 471–492.

Dijk, Oege, Martin Holmen, and Michael Kirchler, "Rank matters—The impact of social competition on portfolio choice," *European Economic Review*, 2014, *66*, 97–110.

Ding, Bill, Mila Getmansky, Bing Liang, and Russ Wermers, "Share restrictions and investor flows in the hedge fund industry," *SSRN eLibrary*, 2009, *111.*

Drechsler, Itamar, "Risk choice under high-water marks," *Review of Financial Studies*, 2014, *27* (7), 2052–2096.

Dyck, Alexander, Karl V Lins, and Lukasz Pomorski, "Does active management pay? New international evidence," *Review of Asset Pricing Studies*, 2013, *3* (2), 200–228.

Edelen, Roger M, "Investor flows and the assessed performance of open-end mutual funds," *Journal of Financial Economics*, 1999, *53* (3), 439–466.

Ewerhart, Christian, "Mixed equilibria in Tullock contests," *Economic Theory*, 2015, *60* (1), 59–71.

_ , "Contests with small noise and the robustness of the all-pay auction," *Games and Economic Behavior*, 2017, *105*, 195–211.

Fama, Eugene F and Kenneth R French, "Luck versus skill in the cross-section of mutual fund returns," *The Journal of Finance*, 2010, *65* (5), 1915–1947.

Fang, Dawei and Thomas H Noe, "Skewing the odds: Strategic risk taking in contests," *Available at SSRN 2339212*, 2015.

Feng, Han and David Hobson, "Gambling in contests with regret," *Mathematical Finance*, 2014, *Article first published online: 20 MAY 2014. DOI: 10.1111/mafi.12069.*

_ , _ et al., "Gambling in contests with random initial law," *The Annals of Applied Probability*, 2016, *26* (1), 186–215.

Feroli, Michael, Anil K Kashyap, Kermit L Schoenholtz, and Hyun Song Shin, "Market tantrums and monetary policy," *Chicago Booth Research Paper*, 2014, (14-09).

French, Kenneth R, "Presidential address: The cost of active investing," *The Journal of Finance*, 2008, *63* (4), 1537–1573.

Friend, Irwin and Marshall E Blume, "The demand for risky assets," *American Economic Review*, 1975, pp. 900–922.

Frydman, Carola and Raven E Saks, "Executive compensation: A new view from a long-term perspective, 1936–2005," *Review of Financial Studies*, 2010, *23* (5), 2099–2138.

Fu, Qiang and Jingfeng Lu, "Micro foundations of multi-prize lottery contests: A perspective of noisy performance ranking," *Social Choice and Welfare*, 2012, *38* (3), 497–517.

Garcia, Rene, Richard Luger, and Eric Renault, "Empirical assessment of an intertemporal option pricing model with latent variables," *Journal of Econometrics*, 2003, *116* (1-2), 49–83.

Gârleanu, Nicolae and Lasse Heje Pedersen, "Efficiently inefficient markets for assets and asset management," *The Journal of Finance*, 2018, *73* (4), 1663–1712.

_ **and** _ , "Active and Passive Investing," *Available at SSRN 3253537*, 2019.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, "Money doctors," *The Journal of Finance*, 2015, *70* (1), 91–114.

_ , _ , and Robert W Vishny, "A model of shadow banking," *The Journal of Finance*, 2013, *68* (4), 1331–1363.

Gentzkow, Matthew and Emir Kamenica, "Costly persuasion," *American Economic Review*, 2014, *104*, 457–462.

Glasserman, Paul and Harry Mamaysky, "Investor information choice with macro and micro information," *Columbia Business School Research Paper*, 2018, (16-45).

Glosten, Lawrence R, Suresh Nallareddy, and Yuan Zou, "ETF activity and informational efficiency of underlying securities," *Columbia Business School Research Paper*, 2016, (16-71).

Goel, Anand M. and Anjan V. Thakor, "Overconfidence, CEO selection, and corporate governance," *The Journal of Finance*, 2008, *63* (6), 2737–2784.

Goetzmann, William N, Jonathan E Ingersoll, and Stephen A Ross, "High-water marks and hedge fund management contracts," *The Journal of Finance*, 2003, *58* (4), 1685–1718.

Goldstein, Itay and Ady Pauzner, "Demand–deposit contracts and the probability of bank runs," *The Journal of Finance*, 2005, *60* (3), 1293–1327.

_ , Hao Jiang, and David T Ng, "Investor flows and fragility in corporate bond funds," *Journal of Financial Economics*, 2017, *126* (3), 592–613.

Gomes, Francisco and Alexander Michaelides, "Asset pricing with limited risk sharing and heterogeneous agents," *Review of Financial Studies*, 2007, *21* (1), 415–448.

Gomez, Matthieu et al., "Asset prices and wealth inequality," *Unpublished paper: Princeton*, 2019.

Green, Richard C, "Investment incentives, debt, and warrants," *Journal of Financial Economics*, 1984, *13* (1), 115–136.

Greenwood, Robin, "Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights," *Review of Financial Studies*, 2007, *21* (3), 1153–1186.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa, "IQ and stock market participation," *The Journal of Finance*, 2011, *66* (6), 2121–2164.

Grossman, Sanford J and Joseph E Stiglitz, "On the impossibility of informationally efficient markets," *American Economic Review*, 1980, *70* (3), 393–408.

Guasoni, Paolo and Jan Obłój, "The incentives of hedge fund fees and high-water marks," *Mathematical Finance*, 2016, *26* (2), 269–295.

Guercio, Diane Del and Paula A Tkac, "Star power: The effect of monrningstar ratings on mutual fund flow," *Journal of Financial and Quantitative Analysis*, 2008, *43* (04), 907–936.

Hamm, Sophia JW, "The effect of ETFs on stock liquidity," *Available at SSRN 1687914*, 2014.

Harris, Lawrence and Eitan Gurel, "Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures," *The Journal of Finance*, 1986, *41* (4), 815–829.

Harris, Robert, Tim Jenkinson, and Rüdiger Stucke, "Are too many private equity funds top quartile?," *Journal of Applied Corporate Finance*, 2012, *24*, 77–89.

Haubrich, Joseph G, "Risk aversion, performance pay, and the principal-agent problem," *Journal of Political Economy*, 1994, *102* (2), 258–276.

He, Xue Dong and Steven Kou, "Profit sharing in hedge funds," *Mathematical Finance*, 2018, *28* (1), 50–81.

He, Zhiguo, "Optimal executive compensation when firm size follows geometric brownian motion," *Review of Financial Studies*, 2009, *22* (2), 859–892.

Hillman, Arye L. and Dov Samet, "Dissipation of contestable rents by small numbers of contenders," *Public Choice*, 1987, *54*, 63–82.

Hillman, Arye L and John G Riley, "Politically contestable rents and transfers," *Economics & Politics*, 1989, *1* (1), 17–39.

Hodder, James E and Jens Carsten Jackwerth, "Incentive contracts and hedge fund management," *Journal of Financial and Quantitative Analysis*, 2007, *42* (04), 811–826.

Holland, John L, *Making Vocational Choices: A Theory of Careers*, Englewood Cliffs, 1973.

Holmstrom, Bengt and Paul Milgrom, "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica: Journal of the Econometric Society*, 1987, pp. 303–328.

_ and _, "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," *Journal of Law, Economics, & Organization*, 1991, *7*, 24–52.

Hong, Harrison, Jeffrey D Kubik, and Jeremy C Stein, "Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers," *The Journal of Finance*, 2005, *60* (6), 2801–2824.

Huang, Jennifer, Kelsey D Wei, and Hong Yan, "Participation costs and the sensitivity of fund flows to past performance," *The Journal of Finance*, 2007, *62* (3), 1273–1311.

Israeli, Doron, Charles MC Lee, and Suhas A Sridharan, "Is there a dark side to exchange traded funds? An information perspective," *Review of Accounting Studies*, 2017, *22* (3), 1048–1083.

Iyer, Rajkamal and Manju Puri, "Understanding bank runs: The importance of depositor-bank relationships and networks," *American Economic Review*, 2012, *102* (4), 1414–45.

James, Kevin R., Daniel Mittendorf, Andrea Pirrone, and Claudia Robles-Garcia, "Does the growth of passive investing affect equity market performance?: A literature review," *FCA Occasional Paper*, 2019.

Janeček, Karel and Mihai Sîrbu, "Optimal investment with high-watermark performance fee," *SIAM Journal on Control and Optimization*, 2012, *50* (2), 790–819.

Jensen, Michael C, "The performance of mutual funds in the period 1945–1964," *The Journal of Finance*, 1968, *23* (2), 389–416.

_ **and Kevin J Murphy**, "Performance pay and top-management incentives," *Journal of Political Economy*, 1990, *98* (2), 225–264.

Johnson, Woodrow T, "Predictable investment horizons and wealth transfers among mutual fund shareholders," *The Journal of Finance*, 2004, *59* (5), 1979–2012.

Jr, Henry L Tosi and Luis R Gomez-Mejia, "The decoupling of CEO pay and performance: An agency theory perspective," *Administrative Science Quarterly*, 1989, pp. 169–189.

Kacperczyk, Marcin and Amit Seru, "Fund manager use of public information: New evidence on managerial skills," *The Journal of Finance*, 2007, *62* (2), 485–528.

_ **and Philipp Schnabl**, "How safe are money market funds?," *The Quarterly Journal of Economics*, 2013, *128* (3), 1073–1122.

_ **, Clemens Sialm, and Lu Zheng**, "On the industry concentration of actively managed equity mutual funds," *The Journal of Finance*, 2005, *60* (4), 1983–2011.

_ **, Stijn Van Nieuwerburgh, and Laura Veldkamp**, "Time-varying fund manager skill," *The Journal of Finance*, 2014, *69* (4), 1455–1484.

_ **, Stijn Van Nieuwerburgh, and Laura Veldkamp**, "A rational theory of mutual funds' attention allocation," *Econometrica*, 2016, *84* (2), 571–626.

Kale, Jayant R, Ebru Reis, and Anand Venkateswaran, "Rank-order tournaments and incentive alignment: The effect on firm performance," *The Journal of Finance*, 2009, *64* (3), 1479–1512.

Kamenica, Emir and Matthew Gentzkow, "Bayesian persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

Kaplan, Steven N and Antoinette Schoar, "Private equity performance: Returns, persistence, and capital flows," *The Journal of Finance*, 2005, *60* (4), 1791–1823.

Kay, Ira, Steve Friedman, Brian Lane, Blaine Martin, and Soren Meischeid, "Are companies setting challenging target incentive goals?," https://corpgov.law.harvard.edu/2015/04/18/are-companies-setting-challenging-target-incentive-goals/ 2015.

Kempf, Alexander and Stefan Ruenzi, "Tournaments in mutual-fund families," *Review of Financial Studies*, 2008, *21* (2), 1013–1036.

Kini, Omesh and Ryan Williams, "Tournament incentives, firm risk, and corporate policies," *Journal of Financial Economics*, 2012, *103* (2), 350–376.

Kirchler, Michael, Florian Lindner, and Utz Weitzel, "Rankings and risk-taking in the finance industry," *The Journal of Finance*, 2018, *73* (5), 2271–2302.

Koijen, Ralph SJ, Robert J Richmond, and Motohiro Yogo, "Which investors matter for global equity valuations and expected returns?," *Available at SSRN*, 2019.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, "Can mutual fund 'stars' really pick stocks? New evidence from a bootstrap analysis," *The Journal of Finance*, 2006, *61* (6), 2551–2595.

Krasny, Yoel, "Asset pricing with status risk," *The Quarterly Journal of Finance*, 2011, *1* (03), 495–549.

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

Lambert, Peter J and J Richard Aronson, "Inequality decomposition analysis and the Gini coefficient revisited," *Economic Journal*, 1993, *103* (420), 1221–1227.

Lan, Yingcong, Neng Wang, and Jinqiang Yang, "The economics of hedge funds," *Journal of Financial Economics*, 2013, *110* (2), 300–323.

Lauriola, Marco and Irwin P Levin, "Personality traits and risky decision-making in a controlled experimental task: An exploratory study," *Personality and Individual Differences*, 2001, *31* (2), 215–226.

Lazear, Edward P, "Pay equality and industrial politics," *Journal of Political Economy*, 1989, *97* (3), 561–580.

_ **and Sherwin Rosen**, "Rank-order tournaments as optimum labor contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

Lee, Jung Hoon, Charles Trzcinka, and Shyam Venkatesan, "Mutual fund risk-shifting and management contracts," *Available at SSRN*, 2016.

_ , _ , **and** _ , "Do portfolio manager contracts contract portfolio management?," *The Journal of Finance*, 2019, *74* (5), 2543–2577.

Levine, David I, "Cohesiveness, productivity, and wage dispersion," *Journal of Economic Behavior & Organization*, 1991, *15* (2), 237–255.

Lewellen, Jonathan, "Institutional investors and the limits of arbitrage," *Journal of Financial Economics*, 2011, *102* (1), 62–80.

Lewrick, Ulf and Jochen F Schanz, "Is the price right? Swing pricing and investor redemptions," 2017.

_ **and** _ , "Liquidity risk in markets with trading frictions: What can swing pricing achieve?," 2017.

Lim, Jongha, Berk A Sensoy, and Michael S Weisbach, "Indirect incentives of hedge fund managers," *The Journal of Finance*, 2015.

Lindner, Florian, Michael Kirchler, Stephanie Rosenkranz, and Utz Weitzel, "Social status and risk-taking in investment decisions," 2019. MPI Collective Goods Discussion Paper.

Liu, Xuewen and Antonio S Mello, "The fragile capital structure of hedge funds and the limits to arbitrage," *Journal of Financial Economics*, 2011, *102* (3), 491–506.

Loughran, Tim and Jay R Ritter, "The operating performance of firms conducting seasoned equity offerings," *The Journal of Finance*, 1997, *52* (5), 1823–1850.

Luenberger, David, *Optimization by Vector Space Methods*, Wiley, 1969.

Ma, Linlin and Yuehua Tang, "Portfolio manager ownership and mutual fund risk taking," *Management Science*, 2019, *65* (12), 5518–5534.

_ , _ , **and Juan pedro Gómez**, "Portfolio manager compensation in the US mutual fund industry," *The Journal of Finance*, 2019, *74* (2), 587–638.

Makarov, Dmitry and Astrid V Schornick, "A note on wealth effect under CARA utility," *Finance Research Letters*, 2010, *7* (3), 170–177.

Malamud, Semyon, "A dynamic equilibrium model of ETFs," 2016.

Malikov, George, "Information, participation, and passive investing," *Available at SSRN*, 2019.

Mamaysky, Harry, Matthew Spiegel, and Hong Zhang, "Estimating the dynamics of mutual fund alphas and betas," *Review of Financial Studies*, 2008, *21* (1), 233–264.

Mankiw, NG, "Consumer durables and the real interest rate," *Review of Economics and Statistics*, 1985, *67* (3), 353–362.

Martynova, Natalya and Enrico Perotti, "Convertible bonds and bank risk-taking," *Journal of Financial Intermediation*, 2018, *35*, 61–80.

Matějka, Michal and Korok Ray, "Balancing difficulty of performance targets: Theory and evidence," *Review of Accounting Studies*, 2017, *22* (4), 1666–1697.

Merton, Robert C, "Lifetime portfolio selection under uncertainty: The continuous-time case," *Review of Economics and Statistics*, 1969, pp. 247–257.

Metrick, Andrew and Ayako Yasuda, "The economics of private equity funds," *Review of Financial Studies*, 2010, *23* (6), 2303–2341.

Mitchell, Mark, Lasse Heje Pedersen, and Todd Pulvino, "Slow moving capital," *American Economic Review*, 2007, *97* (2), 215–220.

Morris, Stephen and Hyun Song Shin, "Unique equilibrium in a model of self-fulfilling currency attacks," *American Economic Review*, 1998, pp. 587–597.

_ , Ilhyock Shim, and Hyun Song Shin, "Redemption risk and cash hoarding by asset managers," *Journal of Monetary Economics*, 2017, *89*, 71–87.

Murphy, Kevin J, "Executive compensation," *Handbook of Labor Economics*, 1999, *3*, 2485–2563.

Myerson, Roger B., "Incentives to cultivate favored minorities under alternative electoral systems," *American Political Science Review*, 1993, *87* (04), 856–869.

Noe, Thomas, "Comparing the chosen: Selection bias when selection is competitive," *Journal of Political Economy*, 2020, *128* (1), 000–000.

Olszewski, Wojciech and Ron Siegel, "Effort-maximizing contests," Technical Report, mimeo 2016.

_ and _ , "Large contests," *Econometrica*, 2016, *84* (2), 835–854.

Oppel, RA, "The index monster in your closet," *The New York Times*, 1999, p. 17.

Palomino, Frédéric and Andrea Prat, "Risk taking and optimal contracts for money managers," *RAND Journal of Economics*, 2003, *34* (1), 113–137.

Panageas, Stavros and Mark M Westerfield, "High-water marks: High risk appetites? Convex compensation, long horizons, and portfolio choice," *The Journal of Finance*, 2009, *64* (1), 1–36.

Pástor, L'uboš and Robert F Stambaugh, "On the size of the active management industry," *Journal of Political Economy*, 2012, *120* (4), 740–781.

_ , _ , and Lucian A Taylor, "Do funds make more when they trade more?," *The Journal of Finance*, 2017, *72* (4), 1483–1528.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski, "Responsible investing: The ESG-efficient frontier," *Available at SSRN 3466417*, 2019.

Peng, Lin and Wei Xiong, "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, 2006, *80* (3), 563–602.

Peress, Joel, "Wealth, information acquisition, and portfolio choice," *Review of Financial Studies*, 2003, *17* (3), 879–914.

Petajisto, Antti, "Active share and mutual fund performance," *Financial Analysts Journal*, 2013, *69* (4), 73–93.

Phalippou, Ludovic, Christian Rauch, and Marc Umber, "Private equity portfolio company fees," *Journal of Financial Economics*, 2018, *129* (3), 559–585.

Proto, Eugenio, Aldo Rustichini, and Andis Sofianos, "Intelligence, personality, and gains from cooperation in repeated interactions," *Journal of Political Economy*, 2019, *127* (3), 1351–1390.

Ramaswamy, Srichander, "Market structures and systemic risks of exchange-traded funds," 2011.

Ravid, S Abraham and Matthew Spiegel, "Optimal financial contracts for a start-up with unlimited operating discretion," *Journal of Financial and Quantitative Analysis*, 1997, *32* (3), 269–286.

Ray, Debraj and Arthur Robson, "Status, intertemporal choice, and risk-taking," *Econometrica*, 2012, *80* (4), 1505–1531.

Rees, E. L., "Graphical discussion of the roots of a quartic equation," *American Mathematical Monthly*, 1922, *29* (2), 51–55.

Rindermann, Heiner and James Thompson, "Cognitive capitalism: The effect of cognitive ability on wealth, as mediated through scientific achievement and economic freedom," *Psychological Science*, 2011, *22* (6), 754–763.

Robson, Arthur J., "Status, the distribution of wealth, private and social attitudes to risk," *Econometrica*, 1992, *60* (4), 837–857.

Rock, Edward B and Marcel Kahan, "Index funds and corporate governance: Let shareholders be shareholders," *NYU Law and Economics Research Paper*, 2018, (18-39).

Rose-Ackerman, Susan, "Risk taking and ruin: Bankruptcy and investment choice," *The Journal of Legal Studies*, 1991, *20* (2), 277–310.

Ross, Stephen A, "Compensation, incentives, and the duality of risk aversion and riskiness," *The Journal of Finance*, 2004, *59* (1), 207–225.

Roussanov, Nikolai, "Diversification and its discontents: Idiosyncratic and entrepreneurial risk in the quest for social status," *The Journal of Finance*, 2010, *65* (5), 1755–1788.

Ruef, Martin, Howard E. Aldrich, and Nancy M. Carter, "The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs," *American Sociological Review*, 2003, *68* (2), 195–222.

Saez, Emmanuel and Gabriel Zucman, "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data," *The Quarterly Journal of Economics*, 2016, *131* (2), 519–578.

Saha, Atanu, "Expo-power utility: A 'flexible' form for absolute and relative risk aversion," *American Journal of Agricultural Economics*, 1993, *75* (4), 905–913.

Sannikov, Yuliy, "A continuous-time version of the principal-agent problem," *Review of Economic Studies*, 2008, *75* (3), 957–984.

Schmidt, Lawrence, Allan Timmermann, and Russ Wermers, "Runs on money market mutual funds," *American Economic Review*, 2016, *106* (9), 2625–57.

Schöttner, Anja, "Relational contracts, multitasking, and job design," *The Journal of Law, Economics, & Organization*, 2007, *24* (1), 138–162.

Seel, Christian and Philipp Strack, "Gambling in contests," *Journal of Economic Theory*, 2013, *148* (5), 2033–2048.

_ and _, "Continuous time contests with private information," *Mathematics of Operations Research*, 2016.

Sharpe, William F, "The arithmetic of active management," *Financial Analysts Journal*, 1991, *47* (1), 7–9.

Siegel, Ron, "All-pay contests," *Econometrica*, 2009, *77* (1), 71–92.

_ , "Asymmetric contests with conditional investments," *American Economic Review*, 2010, *100* (5), 2230–2260.

Simon, Leo K and William R Zame, "Discontinuous games and endogenous sharing rules," *Econometrica*, 1990, *58* (4), 861–872.

Sorensen, Eric H, Keith L Miller, and Vele Samak, "Allocating between active and passive management," *Financial Analysts Journal*, 1998, *54* (5), 18–31.

Sorensen, Morten, Neng Wang, and Jinqiang Yang, "Valuing private equity," *Review of Financial Studies*, 2014, *27* (7), 1977–2021.

Staer, Arsenio, "Equivalent volume and comovement," *Available at SSRN 2158462*, 2014.

Starks, Laura T, "Performance incentive fees: An agency theoretic approach," *Journal of Financial and Quantitative Analysis*, 1987, *22* (01), 17–32.

Strack, Philipp, "Risk-taking in contests: The impact of fund-manager compensation on investor welfare," *Available at SSRN 2739177*, 2016.

Sushko, Vladyslav and Grant Turner, "The implications of passive investing for securities markets," *BIS Quarterly Review, March*, 2018.

Tosi, Henry L, Steve Werner, Jeffrey P Katz, and Luis R Gomez-Mejia, "How much does performance matter? A meta-analysis of CEO pay studies," *Journal of Management*, 2000, *26* (2), 301–339.

Van Wesep, Edward D and Sean Wang, "The prevention of excess managerial risk taking," *Journal of Corporate Finance*, 2014, *29*, 579–593.

Vayanos, Dimitri and Paul Woolley, "Curse of the benchmarks," *Paul Woolley Center at the London School of Economics Discussion Paper# 747*, 2016.

Vijh, Anand M, "S&P 500 trading strategies and stock betas," *Review of Financial Studies*, 1994, *7* (1), 215–251.

Vives, Xavier, "Strategic complementarity, fragility, and regulation," *Review of Financial Studies*, 2014, *27* (12), 3547–3592.

Xiao, Jun, "Asymmetric all-pay contests with heterogeneous prizes," *Journal of Economic Theory*, 2016, *163*, 178–221.

_ , "Equilibrium analysis of the all-pay contest with two nonidentical prizes: Complete results," *Journal of Mathematical Economics*, 2018, *74*, 21–34.

Yitzhaki, Shlomo, "Relative deprivation and the Gini coefficient," *The Quarterly Journal of Economics*, 1979, pp. 321–324.

Zagorsky, Jay L, "Do you have to be smart to be rich? The impact of IQ on wealth, income and financial distress," *Intelligence*, 2007, *35* (5), 489–501.

Zeng, Yao, "A dynamic theory of mutual fund runs and liquidity management," *Available at SSRN 2907718*, 2017.