Essays on Mathematical Finance

Georgios Vichos

Department of Statistics
London School of Economics and Political Science

This dissertation is submitted for the degree of
Doctor of Philosophy

April 2019

I would like to dedicate this thesis to my loving parents.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of less than 100,000 words.

I confirm that Chapter 2 was jointly co-authored with Professor Michail Antrhopelos and Professor Konstantinos Kardaras and I contributed 33% of this work.

I confirm that Chapter 3 was jointly co-authored with Dimitrios Papadimitriou and Konstantinos Tokis and I contributed 33% of this work.

Georgios Vichos
April 2019

Acknowledgements

I would like to thank my advisor Konstantinos Kardaras for his support and help, both scientifically and morally during my academic trip.

I would also like to thank Dimitris Papadimitriou, Konstantinos Tokis, Emmanouil Androulakis, Dimitris Chatzakos, Nick Savvidis, Diego Zabaljauregui, Stavros Kevopoulos and Ian Marshall for helpful comments, discussions and encouragement.

I would also like to thank Konstantinos Kalogeropoulos, Beatrice Acciaio, Michail Anthropelos, Dimitri Vayanos, Johannes Ruf, Jonathan Berk, Amil Dasgupta, Peter Kondor and Dong Lou, for discussions that inspired this thesis.

This thesis would not have being possible without the generous financial support of the EPSRC and Philip Treleaven.

Most importantly I devoutly would like to thank Pafuni for her unconditional love, support and care throughout the volatile PhD journey.

Abstract

The first part of this thesis deals with the consideration of thin incomplete financial markets, where traders with heterogeneous preferences and risk exposures have motive to behave strategically regarding the demand schedules they submit, thereby impacting prices and allocations. We argue that traders relatively more exposed to market risk tend to submit more elastic demand functions. Noncompetitive equilibrium prices and allocations result as an outcome of a game among traders. General sufficient conditions for existence and uniqueness of such equilibrium are provided, with an extensive analysis of two-trader transactions. Even though strategic behaviour causes inefficient social allocations, traders with sufficiently high risk tolerance and/or large initial exposure to market risk obtain more utility gain in the noncompetitive equilibrium, when compared to the competitive one.

The second part of this thesis considers a continuum of potential investors allocating funds in two consecutive periods between a manager and a market index. The manager's alpha, defined as her ability to generate idiosyncratic returns, is her private information and is either high or low. In each period, the manager receives a private signal on the potential performance of her alpha, and she also obtains some public news on the market's condition. The investors observe her decision to either follow a market neutral strategy, or an index tracking one. It is shown that the latter always results in a loss of reputation, which is also reflected on the fund's flows. This loss is smaller in bull markets, when investors expect more managers to use high beta strategies. As a result, a manager's performance in bull markets is less informative about her ability than in bear markets, because a high beta strategy does not rely on it. We empirically verify that flows of funds that follow high beta strategies are less responsive to the fund's performance than those that follow market neutral strategies.

Contents

List of Tables

Chapter 1

Introduction

1.1 Thesis outline

The thesis is organised as follows:

Chapter 1 gives an overview and the contribution of the thesis.

Chapter 2 is about ”Effective risk aversion in risk sharing games”. We consider thin incomplete financial markets, where traders with heterogeneous preferences and risk exposures have motive to behave strategically regarding the demand schedules they submit, thereby impacting prices and allocations. We argue that traders relatively more exposed to market risk tend to submit more elastic demand functions. Noncompetitive equilibrium prices and allocations result as an outcome of a game among traders. General sufficient conditions for existence and uniqueness of such equilibrium are provided, with an extensive analysis of two-trader transactions. Even though strategic behaviour causes inefficient social allocations, traders with sufficiently high risk tolerance and/or large initial exposure to market risk obtain more utility gain in the noncompetitive equilibrium, when compared to the competitive one.

Chapter 3 deals with ”The effect of market conditions and career concerns in the fund industry”. A continuum of potential investors allocate funds in two consecutive periods between a manager and a market index. The manager’s alpha, defined as her ability to generate idiosyncratic returns, is her private information and is either high or low. In each period, the manager receives a private signal on the potential performance of her alpha, and she also obtains some public news on the market’s condition. The investors observe her decision to either follow a market neutral strategy, or an index tracking one. It is shown that the latter always results in a loss of reputation, which is also reflected on the fund’s flows.

This loss is smaller in bull markets, when investors expect more managers to use high beta strategies. As a result, a manager's performance in bull markets is less informative about her ability than in bear markets, because a high beta strategy does not rely on it. We empirically verify that flows of funds that follow high beta strategies are less responsive to the fund's performance than those that follow market neutral strategies.

1.2 Major contributions

The purpose of this thesis is to contribute to two topics in financial mathematics. Firstly, a risk sharing game is studied in a possibly incomplete thin financial market, where investors strategically declare risk tolerances—not necessarily reflecting their risky profile—in order to share their risk exposures by means of trading. Secondly, it proposes an equilibrium model to facilitate how market conditions and concerns of fund managers about their reputation, impact the fund industry and consequently empirically validates some of the theoretical findings in this framework. More specifically:

- The domination of many financial markets by a small number of large investors and the influence of price and allocation of tradeable assets has been widely recognised; see among others, Blume and Keim (2012), Gibson et al. (2003), Rostek and Weretka (2015). This type of market impact has been observed in large exchanges like NYSE (see Keim and Madhavan (1995, 1996), Madhavan and Cheng (1997), Hameed et al. (2017)), especially in over the counter transactions that the assumption of a competitive market structure cannot hold. In reality the majority of these types of transactions involve a few participants and although all the information is public, equilibrium forms in a game-theoretic manner.

  Chapter 2's purpose is to model imperfect competition between traders by assuming that they submit linear net demand functions with a slope that might be different from their competitive demand function. The random endowments of the investors and the terminal dividends of the traded securities are assumed to be Gaussian. The slope of the trader's demand function depends on their risk tolerance. When these risk tolerances are considered public information, explicit formulas for the competitive equilibrium allocations and the prices of securities can be obtained. While the agents' risk exposures may be public information, their risk aversion can be regarded as private information. Motivated by these observations, the chapter studies noncompetitive equilibrium problem arising when investors act strategically by submitting demand functions with an elasticity that does not necessarily reflect their risk aversion. In this risk sharing game, the study provides explicit formulas for the best response functions in terms of the traders' pre transaction projected betas on risk exposures. These formulas are then used to prove the existence and uniqueness of the Nash equilibrium. More precisely, extreme noncompetitive equilibria when an agent acts as risk neutral are completely characterised. Additionally, if at most one agent is highly exposed to market risk, then a Nash equilibrium exists and is unique. In the bilateral (two-agent) risk sharing game, explicit formulas for the noncompetitive equilibrium are given.

These formulas facilitate the analysis of individual and aggregate utility gains and losses, as well as risk sharing inefficiencies, when compared to the outcome under the competitive equilibrium setup or a complete market. One of the main conclusions of the analysis is that, even though aggregate utility decreases, investors with sufficiently high risk exposures or sufficient high risk tolerance might benefit from the game. These gains are reduced by the incompleteness of the market. Furthermore, the game leads to risk sharing inefficiencies as the post transaction betas are higher than those obtained in the competitive set up.

- The concern of the role of financial intermediaries such as hedge funds and mutual finds has been growing alongside the proportion of the institutional ownership of equities. The competition of the managers and their concern about their reputation may influence their investment decisions in a way that is not necessarily optimal. One of the seminal papers about mutual funds is Berk and Green (2004), where the lack of persistence of outperformance is explained by the competition between funds and reallocation of investors' capital and not by the lack of managers' skills. The connection of reputation and investment decisions has been recently studied and one can remark the following: risk taking behaviour in order to increase manager's reputation leads to overinvestment (Chen (2015)), and reputation concerns lead to herding and some other anomalies and career concerns create a reputational premium which depends on the economic conditions (Dasgupta and Prat (2008)). Moreover, Malliaris and Yan (2015) show that career concerns induce a preference over the skewness of their strategy returns, while Hu et al. (2011) present a model of fund industry in which managers alter their risk-taking behaviour based on their past performance; however, they do not take into account any strategic behaviour by the fund managers.

  Chapter 3 is motivated by the interaction between concerns about reputation and investment decisions of fund managers. The study proposes a two period equilibrium model where the manager's investment decisions provide imperfect information about their managing abilities. More precisely, investors receive initially a shock that will determine their wealth allocation for the first period. After the investors' decision, the manager receives a private signal on her idiosyncratic strategy and a public signal on market conditions. She then decides whether to invest in the market or in her idiosyncratic strategy. After observing the returns and the actions of the manager in the first period, investors update their beliefs about the ability of the manager. Based on this Bayesian update of reputation, they allocate their wealth for the second and final period. In this set up, it is clear that the manager's actions are influenced by the consequences that they could have on her reputation after the first period, and hence on the assets that she will have under management on the second period. The analysis then focuses on monotonic equilibrium, these are equilibria where the manager's reputation is a nondecreasing function of her performance.

- As a first result, a refinement of the perfect Bayesian Equilibrium is analysed, which is called *monotonic equilibrium*, and it is shown that this always exists. The only additional restriction that this refinement is imposing is that the manager's reputation is non-decreasing on her performance. In addition, under mild parametric restrictions it is demonstrated that the monotonic equilibrium is unique.
- As a second result, it is demonstrated that investing in an idiosyncratic strategy carries a reputational benefit. This is because the cut-off of the high manager type is smaller than that of the low manager types. In other words, the high type is more receptive to the idea of adopting a low beta strategy. Intuitively, the manager's choice is affected by two incentives. On the one hand, she wants to increase her reputation, which skews her preferences towards idiosyncratic investments. On the other hand, she cares about the realised return of her strategy, since her fees depend on it. Hence, for a relatively low private signal, even a high type may opt to forfeit the reputational benefit, because investing in the market will generate higher returns, and as a result more fees. Therefore, the investment strategy is informative but does not fully reveal the manager's ability, which is a realistic representation of the fund industry.
- Finally, as the third and most important result, it is shown that the reputational benefit of investing in the idiosyncratic project is decreasing in the market conditions. In particular, it is proven that the expected sensitivity of reputation to performance is higher in bear markets than in bull markets. This is because investors understand the dual objective of managers and the fact that a manager is more likely to invest in the market when the market conditions are good, and thus update their beliefs less aggressively when this is the case; instead, in bad times any change in a fund's performance is much more likely to be attributed to the ability of the manager.
- The above results are used to discuss the competition between funds, in terms of their sizes, and its fluctuation depending on market conditions. It is predicted that the likelihood of changes in the ranking of the funds, measured by assets under management, is hump-shaped on the market return, but is also higher during bear markets than during bull markets, due to the higher informativeness of performance. Some empirical evidence is found that supports this prediction.

This is in line with the common perception that the industry only rearranges its interaction with its investors during crises.

Some of the assumptions and findings from the theoretical model are confirmed by an empirical analysis using data from the Morningstar CISDM. At the end of the chapter it is shown that it is impossible to find monotonic equilibria if the betas of the managers are unobservable.

Chapter 2

Effective Risk Aversion In Thin Risk-Sharing Markets

Introduction

It has been widely recognised that many financial markets are dominated by a relatively small number of large investors, whose actions heavily influence prices and allocations of tradeable securities—see, among others, discussions in Blume and Keim (2012), Gibson et al. (2003), Rostek and Weretka (2015). While such market impact has been observed even in large exchanges like NYSE (see Keim and Madhavan (1995, 1996), Madhavan and Cheng (1997) and the more recent empirical study Hameed et al. (2017)), it is especially in over-the-counter (OTC) transactions that the assumption of a competitive market structure is problematic. The majority of OTC markets involve relatively few participants; therefore, even if all information is public, equilibrium forms in a noncompetitive manner. Such financial markets with an oligopolistic structure are usually characterised as *thin* (see Rostek and Weretka (2008) for a related reviewing discussion).

The main reason for trading between risk averse traders with common information and beliefs is the heterogeneity of their endowments—see, for instance, related discussion in Barrieu and El Karoui (2005), Jouini et al. (2008). Trading securities that are correlated with traders' endowments may be mutually beneficial in sharing the traders' risky positions—see, among others, Anthropelos and Žitković (2010), Robertson (2017). In a standard Walrasian uniform-price auction model, traders submit demand schedules on the tradeable securities and the market clears at the prices resulting in zero aggregate submitted demand; and since demand depends on traders' characteristics such as their risk exposure and risk aversion, the same is true for the equilibrium prices and allocation. Whereas traders' exposures to market risk (i.e., uncertainty in the tradeable securities) may be considered public knowledge, their risk aversion is *subjective* and should be regarded as private information. In the realm of thin financial markets, traders may have motive to act strategically and submit demand schedules with *different* elasticity than the one reflecting their risk aversion. The goal of this paper is to model such strategic behaviour and highlight some of its economic insights.

Model description and main contributions

We develop a model of a one-shot transaction on a given collection of risky tradeable securities, under common information on the probabilistic nature of their payoffs. Traders possess and exploit a potential to impact the market's equilibrium. We adopt the setting of CARA preferences and normally distributed payoffs, also appearing in Kyle (1989), Rostek and Weretka (2015), Vayanos (1999), with traders assumed heterogeneous with respect to risk tolerance (defined as the reciprocal of risk aversion) and initial risky positions. In contrast to the majority of related literature, we do not assume that traders' endowments belong to the span of the tradeable securities, leading to market incompleteness.

Similarly to the models in Kyle (1989), Vayanos (1999), Vives (2011), the market operates as a uniform-price auction where traders submit demand functions on the tradeable securities, with equilibrium occurring at the price vector that clears the market. When traders do not act strategically, the market structure is competitive and the equilibrium price-allocation is induced by traders' *true* demand functions. However, as has been pointed out previously, such competitive structure is not suitable for thin markets, and the way traders behave depends on principle on the risk exposure and risk tolerance of their counter-parties. In a CARA-normal setting, demand functions are linear with a downward slope and their elasticities coincide with the traders' risk tolerance. Traders recognise their ability to influence the equilibrium transaction, and may submit demand with different elasticity than the one reflecting their risk tolerance. We formulate a best-response problem, according to which traders submit demand functions aiming at individual utility maximisation, with strategic choices parametrised by the elasticity of the submitted demand. This forms a noncompetitive market scheme, where the Bayesian Nash equilibrium is the fixed point of traders' best responses.

In any non-trivial case, traders have motive to submit demand with different elasticity than their risk tolerance. The main determining factor of traders' best response is their pre-transaction *projected beta*, defined as the beta (in terms of the Capital Asset Pricing Model) of the projection of the trader's risky position onto the linear space generated by the securities. In the special case where the traders' positions belong to the span of tradeable securities projected and actual betas coincide. Following classical literature, traders' projected betas (hereafter simply called betas) measure their exposure to market risk. In terms of risk sharing, we distinguish traders to those who increase or decrease their beta through the transaction.

It is shown that traders submit demand corresponding to higher risk tolerance if and only if they reduce their market risk exposure through trading. The economic insight of this strategic behaviour is simple: traders with relatively higher initial exposure to market risk pay a *risk premium* to their counter-parties in order to reduce their beta. Submitting a more elastic demand has two main effects. Firstly, the post-transaction reduction of beta is smaller, since more elastic demand implies higher relative risk tolerance and hence higher post-transaction exposure to market risk, as the trader appears willing to keep a more risky position. Secondly, the risk premium that is paid is also lower. As it turns out, the effect of premium reduction overtakes the sub-optimal reduction of market risk exposure. In order to obtain intuition on this, consider the impact of the other traders' status on an individual trader's actions. Large pre-transaction beta for a specific trader implies low aggregate beta for other traders. Acting in a more risk tolerant way, by submitting more elastic demand, a trader essentially exploits this low aggregate exposure to market risk of the counter-parties, and in fact decreases the premium that they ask in order to undertake more market risk.

On the contrary, traders who undertake market risk in exchange for a risk premium, i.e., those with low pre-transaction beta, have motive to submit less elastic demand. Not only does such a strategy result in less undertaken market risk, it also takes advantage of the large aggregate counter-parties' beta, increasing the premium received in order to offset their demand.

Continuing this line of argument, traders overexposed to market risk, with pre-transaction beta sufficiently higher than one, tend to behave as risk neutral, even though their actual risk aversion parameter is strictly positive. In such a case, the trader takes over the whole market risk, reducing the post-transaction beta of their position to one. At the same time, the other traders are willing to offset such transaction since it makes their post-transaction beta equal to zero (i.e., becoming market-neutral); for this reason, they reduce the required risk premium. On the other hand, traders with pre-transaction beta less than or equal to -1 submit extremely inelastic demand functions, implying zero risk tolerance, appearing willing to become market neutral. Again, other traders are eager to offset the transaction, since at this regime their aggregate pre-transaction beta is relatively large, and selling market risk is a very effective hedging transaction.

We discover two regimes of noncompetitive equilibrium. When one of the trader's pre-transaction beta is sufficiently large, there exists a unique linear equilibrium which is *extreme*, in the sense that the market-overexposed trader behaves as being risk neutral and at equilibrium undertakes all market risk. Such extreme Nash equilibrium results in market-neutral portfolios for all other traders, while securities are priced in a risk-neutral manner. In any other "*non-extreme*" cases, noncompetitive equilibria solve a coupled system of quadratic equations, which admits a unique solution under the mild—and rather realistic—assumption that at most one of the traders may have pre-transaction beta greater than one. We provide an efficient constructive proof of the latter fact, which can be used to numerically obtain the unique linear equilibrium given an arbitrary number of traders.

The two-trader case is of special interest, mainly because the large majority of risk-sharing transactions are bilateral between large institutions and/or their clients or brokers; related discussions and statistics are provided in Babus and Hu (2016), Babus (2016), D. et al. (2015), Zawadowski (2013), Hendershott and Madhavan (2015). We obtain explicit expressions for two-trader price-allocation noncompetitive equilibria, which allow us to further analyse the model's economic insight. Noncompetitive and competitive equilibria coincide if and only if the competitive equilibrium transaction is null, in that the initial allocation is already Pareto-optimal. In any other case, for both traders the elasticity of submitted demands in such thin markets deviates from the one utilising their risk tolerances. As emphasised above, the crucial factor is the traders' pre-transaction beta. For non-extreme equilibria we have the following synoptic relationship:

true elasticity < *equilibrium elasticity* $\quad\Leftrightarrow\quad$ *post-transaction beta* < *pre-transaction beta*.

Even if traders have common risk tolerance, deviations between their endowment will make them behave heterogeneously. For a trader with higher (resp., lower) beta, who reduces (resp., increases) market risk through the transaction, the equilibrium elasticity reflects more (resp., less) risk tolerance. One could argue, therefore, that in thin financial markets the assumption of effectively homogeneous risk-averse traders is problematic, since it essentially implies that traders ignore their ability to impact the transaction.

In the context of strategic behaviour, equilibrium prices and allocations are generally impacted. In the two-trader case, the volume in noncompetitive equilibrium is always lower than in the competitive one. More precisely, it is shown that the post-transaction beta after Nash equilibrium is—interestingly enough—the midpoint between the trader's pre-transaction beta and the beta after the competitive transaction. This implies a loss of social efficiency, in the sense that the total utility in noncompetitive equilibrium is reduced when compared to the competitive one. However, such loss of total utility does not always transfer to the individual level. In fact, it follows from the analysis of the bilateral game that the noncompetitive equilibrium is beneficial in terms of utility gain for two types of traders: those with sufficiently high pre-transaction beta, and those with sufficiently high risk tolerance. Such findings in noncompetitive markets are consistent with results in Anthropelos (2017) and Anthropelos and Kardaras (2017). (A result in that spirit also appears in Malamud and Rostek (2017); namely it is shown that, when the market is centralised, less risk averse agents have greater price impact.)

As a final point, and as mentioned above, our model allows for incompleteness, and we study its effect in noncompetitive risk-sharing transaction. Based on the two-trader game, we show that traders who benefit from the noncompetitive market setting (i.e., those with high risk tolerance and/or high exposure to market risk) have their utility gains reduced by the fact that endowments are not securitised, highlighting the importance of completeness especially for large traders that prefer thin markets for sharing risk.

Connections with related literature

The present paper contributes to the large literature on imperfectly competitive financial markets. Based on the seminal works on Nash equilibrium in supply/demand functions of Klemperer and Meyer (1989) and Kyle (1989), most models of noncompetitive markets consider strategically acting agents, whose set of choices corresponds to demand schedules submitted to the transaction. Frequently, the departure from competitive structure stems from informational asymmetry; such is the case in Back (1992), Back et al. (2000), Kyle (1989), Kyle et al. (2018), where agents are categorised as informed, uniformed and noisy. Even without existing risky positions, asymmetric information gives rise to mutually beneficial trading opportunities among traders, who submit demand schedules based on the responses of their counter-parties. Another potential source of noncompetitiveness comes via exogenously imposing asymmetry on the bargaining power among market participants. Bilateral OTC transactions between agents with different bargaining power are modelled in Duffie et al. (2007); in Liu and Wang (2016), it is market makers who possess market power and optimally adjust bid-ask spreads based on submitted orders by informed and uniformed investors. (See the references in Liu and Wang (2016) for alternative models of strategic market makers.) Exogenously imposed differences on market power are also present in Brunnermeier and Pedersen (2005), where traders are divided into price-takers and predatory ones, the latter strategically exploiting the liquidity needs of their counter-parties.

In contrast to the above, our model assumes symmetry for traders' market power; non-competitiveness stems solely from the fact there is a small number of acting traders, each of whom can buy or sell the tradeable securities and has the ability to affect the risk-sharing transaction.[1] The market here is assumed to be oligopolistic, without any form of exogenous frictions or asymmetries.

Market models close to ours considered by other authors include Malamud and Rostek (2017), Rostek and Weretka (2015), Vayanos (1999). In Malamud and Rostek (2017), Rostek and Weretka (2015), Vayanos (1999), and similarly to the present work, traders submit demand in a noncompetitive market setting by taking into account the impact of their orders on the equilibrium. The main difference with our demand-game, when compared to the one-shot market of Rostek and Weretka (2015), Vayanos (1999) and the centralised market of Malamud and Rostek (2017), is the set of traders' strategic choices. More precisely, in these works a trader's price impact is identified as the slope of the submitted aggregate demand of the rest of the traders. Traders estimate (correctly at equilibrium) their price impact and respond by submitting demand schedules aiming at maximising their own utility. In particular, the set of strategic choices consists of the slope of the submitted demand, and equilibrium arises as the fixed point of the traders' price impacts. In our model, we keep the linear equilibrium structure of demand functions and parametrise the set of traders' strategical choice to the submitted *elasticity*, and equilibrium is formed simply at the price where aggregate submitted demand is zero. In this way, each trader responds to the whole demand function of other traders, and not just the slope. This is a crucial trading feature motivated by the benefits of risk sharing, since the intercept point of the demand function corresponds to the traders' exposure to market risk (the correlation of traders' endowment with the tradeable assets). The difference becomes pronounced in the very special case of a single tradeable security, where traders' price impacts of Rostek and Weretka (2015) and Malamud and Rostek (2017) can be seen as the reciprocal of their risk aversion. In Rostek and Weretka (2015), the so-called equilibrium *effective risk aversion*—that is, the risk aversion that is reflected by the equilibrium submitted demands—depends only on the number of traders (as well as a couple of other quantities that we do not use in our model: interest rate and number of allowable trades until the end of each trading round). In particular, heterogeneity of initial risky endowments is not addressed: even with different initial positions at each period, traders do not take into account their counter-parties' exposure to market risk. Our demand-game

[1] Symmetric games in an oligopolistic market of goods (rather than securities with stochastic payoffs) have also been studied in the seminal work of Klemperer and Meyer (1989) and in the more recent papers of Vives (2011) and Weretka (2011). The main structural difference between these market models and ours is that players therein (i.e., firms) can take only the seller's side, while the buyer's side (i.e., the demand for the goods) is essentially exogenous. Additionally, the fact that the tradeable asset is a good creates further technical and economic deviations—for instance, the role of risk exposure is essentially played by the cost function, the price can not be negative, etc. The model in Klemperer and Meyer (1989) imposes randomness on demand, whereas Vives (2011) considers random suppliers' cost and private information status. On the other hand, the model of market power in Weretka (2011) is based on the same setting of price impact as in Rostek and Weretka (2015) and Malamud and Rostek (2017).

may be more appropriate for thin risk-sharing transactions, since it endogenously highlights the importance of traders' initial positions for their strategic behaviour.

Another important trait of our model is that it can be applied to the practically important two-trader case, while the models of Rostek and Weretka (2015), Malamud and Rostek (2017) and Vayanos (1999) are ill-posed for bilateral transactions. As already mentioned, bilateral transactions are a significant part of thin market models, since the majority of the OTC risk-sharing transactions consist of only two counter-parties. Existence of a two-agent Bayesian Nash equilibrium exists under mild assumptions in the model of Rostek and Weretka (2012); however, agents there have private valuations on the tradeable securities.

Further to what was pointed out above, our model allows market incompleteness: tradeable securities do not necessarily span the traders' endowments. We are thus able to generalise the discussion on thin markets and deviations of noncompetitive equilibria from competitive ones in the more realistic framework where traders' endowments are neither securitised nor replicable.

Finally, models of thin risk-sharing markets, albeit with a different set of strategic choices, have been considered in Anthropelos (2017) and Anthropelos and Kardaras (2017). In Anthropelos (2017), traders choose the *endowment* submitted for sharing, and a game on agents' linear demand is formed; in contrast with the present paper, agents in Anthropelos (2017) choose the intercept of the demand function instead of its elasticity. In Anthropelos and Kardaras (2017), traders strategically submit *probabilistic beliefs*, and the model is "inefficiently complete", as securities are endogenously designed by heterogeneous traders in order to share their risky endowments.

Structure of the paper

Section 2.1 introduces the market model and competitive equilibrium, where traders do not act strategically. Section 2.2 introduces, solves and discusses the individual trader's best response problem. Noncompetitive equilibrium is introduced in Section 2.3; general conditions ensuring existence and uniqueness of Nash equilibrium are provided in §2.3.2, conditions for the so-called extreme equilibrium are addressed in §2.3.3. The two-trader game is extensively analysed in Section 2.4. The proof of the main Theorem 2.3.4 is presented in Appendix A.1.

2.1 Model Set-Up

We work on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and denote by $\mathbb{L}^0 \equiv \mathbb{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ the class $\mathcal{F}$-measurable random variables, identified modulo $\mathbb{P}$-a.s. equality.

2.1.1 Agents and preferences

We consider a market of $n+1$ economic traders, where $n \in \mathbb{N} = \{1, 2, \ldots\}$; for concreteness, define the index set $I = \{0, \ldots, n\}$. Traders are assumed risk averse and derive utility only from future consumption of a numéraire at the end of a single period, where all uncertainty is resolved. To simplify the analysis we assume that all considered security payoffs are expressed in units of the numéraire, which implies that future deterministic amounts have the same present value for the traders. Each trader $i \in I$ carries a risky future payoff in units of the numéraire, which is called (random) *endowment*, and denoted by E_i. The endowment $E_i \in \mathbb{L}^0$ denotes the existing risky portfolio of trader $i \in I$, and is not necessarily securitised or tradeable. We define the aggregate endowment $E_I := \sum_{i \in I} E_i$, and set $E \equiv (E_i)_{i \in I}$ to be the vector of traders' endowments.

The preference structure of traders is numerically represented by the functionals

$$\mathbb{L}^0 \ni X \mapsto \mathbb{U}_i(X) := -\delta_i \log \mathbb{E}\left[\exp\left(-X/\delta_i\right)\right] \in [-\infty, \infty), \tag{2.1}$$

where $\delta_i \in (0, \infty)$ is the risk tolerance of trader $i \in I$. Note that $\mathbb{U}_i(X)$ corresponds to the certainty equivalent of potential future random outcome X, when trader $i \in I$ has risk preferences with constant absolute risk aversion (CARA) equal to $1/\delta_i$. It is important to point out that functional $\mathbb{U}_i(\cdot)$ also measures wealth in numéraire units and hence can be used for comparison among different traders (and equilibria). We also define the *aggregate risk tolerance* $\delta_I := \sum_{i \in I} \delta_i$, as well as the *relative risk tolerance* $\lambda_i := \delta_i / \delta_I$ of trader $i \in I$. Note that $\lambda_I \equiv \sum_{i \in I} \lambda_i = 1$. Following standard practice, we shall use subscript "$-i$" to denote aggregate quantities of all traders except trader $i \in I$; for example, $\delta_{-i} := \delta_I - \delta_i$ and $\lambda_{-i} := 1 - \lambda_i$, for all $i \in I$.

2.1.2 Securities and demand

In the market there exist a finite number of tradeable securities indexed by the non-empty set K, with payoffs denoted by $S \equiv (S_k)_{k \in K} \in (\mathbb{L}^0)^K$. The demand function Q_i of trader $i \in I$ on the vector S of securities is given by

$$Q_i(p) := \underset{q \in \mathbb{R}^K}{\operatorname{argmax}}\, \mathbb{U}_i(E_i + \langle q, S - p \rangle), \quad p \in \mathbb{R}^K.$$

Here, and in the sequel, $\langle \cdot, \cdot \rangle$ will denote standard inner product on the Euclidean space $\mathbb{R}^K$.

We follow a classic model of standard literature (e.g. Kyle (1989), Rostek and Weretka (2015), Vayanos (1999) and Vives (2011)) and assume that the joint law of (E, S) is Gaussian. Since traders' endowments do not necessarily belong to the span of S, the market is incomplete. Note also that endowments are not assumed independent of S, or independent of each other. Since only securities in random vector S are tradeable, we identify *market risk* with the variance-covariance matrix of S, denoted by

$$C := \mathbb{C}\mathrm{ov}(S, S).$$

In the sequel we will impose the standing assumption that C has full rank. Additionally, for notational convenience, we shall assume that

$$\mathbb{E}[S_k] = 0, \quad \forall k \in K.$$

Due to the cash-invariance of the traders' certainty equivalent, the latter assumption does not entail any loss of generality, as we can normalise tradeable securities to be $S - \mathbb{E}[S]$. Straightforward computations give

$$\begin{aligned}
\mathbb{U}_i\left(E_i + \langle q, S - p \rangle\right) &= -\delta_i \log \mathbb{E}\left[\exp\left(-(E_i + \langle q, S - p \rangle)/\delta_i\right)\right] \\
&= \mathbb{E}\left[E_i\right] - \langle q, p \rangle - \frac{1}{2\delta_i} \mathbb{V}\mathrm{ar}\left[E_i + \langle q, S \rangle\right] \\
&= \mathbb{E}\left[E_i\right] - \frac{1}{2\delta_i} \mathbb{V}\mathrm{ar}\left[E_i\right] - \frac{1}{2\delta_i} \langle q, Cq \rangle - \left\langle q, p + \frac{1}{\delta_i} \mathbb{C}\mathrm{ov}(E_i, S) \right\rangle.
\end{aligned}$$

We also define the following quantities

$$u_i := \mathbb{E}\left[E_i\right] - \frac{1}{2\delta_i} \mathbb{V}\mathrm{ar}\left[E_i\right] \equiv \mathbb{U}_i(E_i),$$

and, for each $i \in I$,

$$a_i := C^{-1}\mathbb{C}\mathrm{ov}(E_i, S), \quad \text{and} \quad a_{-i} := a_I - a_i,$$

where

$$a_I := \sum_{i \in I} a_i.$$

Then, it follows that

$$\mathbb{U}_i\left(E_i + \langle q, S - p\rangle\right) = u_i - \frac{1}{\delta_i}\langle q, Ca_i\rangle - \frac{1}{2\delta_i}\langle q, Cq\rangle - \langle p, q\rangle,$$

from which we readily obtain that the demand function of trader $i \in I$, given by

$$\mathbb{R}^K \ni p \mapsto Q_i(p) = -a_i - \delta_i C^{-1} p, \quad i \in I, \tag{2.2}$$

is downward-sloping linear. The risk tolerance $\delta_i \in (0, \infty)$ could be considered as the *elasticity* of the demand function of trader $i \in I$, with higher δ_i implying more elastic demand. Furthermore, $a_i \in \mathbb{R}^K$ gives the correlation of the tradeable securities with the endowment of trader $i \in I$, and plays the role of the intercept point of the affine demand function (2.2). According to (2.2), when prices of all securities equal zero, the sign of each element of a_i indicates whether trader $i \in I$ has incentive to buy (when negative) or sell (when positive) the corresponding security.

2.1.3 Competitive equilibrium

While our focus will be on noncompetitive equilibrium, we first define competitive equilibrium of our market, to be used and discussed later as a benchmark for comparison, similarly as in Vayanos (1999) and Vives (2011). Trading the securities represented by S without applying any strategic behaviour (i.e., by assuming a *price-taking* mechanism), the traders reach a competitive equilibrium: prices are determined where the traders' aggregate demand equals zero.

Definition 2.1.1 *The vector $\widehat{p} \in \mathbb{R}^K$ is called **competitive equilibrium prices** if*

$$\sum_{i \in I} Q_i(\widehat{p}) = 0.$$

*The corresponding allocation $(\widehat{q}_i)_{i \in I} \in \mathbb{R}^{K \times I}$ defined via $\widehat{q}_i = Q_i(\widehat{p})$ for all $i \in I$ will be called a **competitive equilibrium allocation** associated to (competitive equilibrium) prices $\widehat{p} \in \mathbb{R}^K$.*

Elementary algebra gives the following result.

Proposition 2.1.2 *There exists a unique competitive equilibrium price $\widehat{p}$ given by*

$$\widehat{p} = -\frac{1}{\delta_I} C a_I, \tag{2.3}$$

with associated competitive equilibrium allocations given by

$$\widehat{q_i} = \lambda_i a_I - a_i, \quad i \in I. \tag{2.4}$$

Remark 2.1.3 *For $i \in I$, $D_i := \langle a_i, S \rangle$ is the projection of the endowment E_i onto the linear span of the tradeable security vector $S \equiv (S_k)_{k \in K}$. At competitive equilibrium, the position of trader $i \in I$, net the price paid, is*

$$\langle \widehat{q_i}, S - \widehat{p} \rangle = \langle \lambda_i a_I - a_i, S \rangle + \frac{1}{\delta_I} \langle \lambda_i a_I - a_i, C a_I \rangle = \lambda_i D_I - D_i - \mathbb{E}_{\mathbb{Q}} [\lambda_i D_I - D_i], \quad i \in I.$$

where $\mathbb{Q}$ is given through $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = \exp(-E_I/\delta_I)/\mathbb{E}_{\mathbb{P}}[\exp(-E_I/\delta_I)]$, where $D_I := \sum_{i \in I} D_i$. In the case where the linear span of the securities equals the linear span of the endowments, it holds that $D_i = E_i - \mathbb{E}_{\mathbb{P}}[E_i]$, for all $i \in I$. Then, the competitive equilibrium coincides with the complete-market Arrow-Debreu risk-sharing equilibrium—see, among others, Borch (1962), Buhlmann (1984) or (Magill and Quinzii, 2002, Chapters 2 and 3).

Remark 2.1.4 *A very special—and as shall be discussed, trivial—situation arises when $a_I = 0$, i.e., when $\mathbb{C}\mathrm{ov}(E_I, S_k) = 0$ holds for every $k \in K$, where we recall that $E_I := \sum_{i \in I} E_i$. In words, $a_I = 0$ means that the total endowment E_I is independent of the spanned subspace of the securities. In this case, in the setting of Proposition 2.1.2, competitive equilibrium prices of the securities are zero, and $\widehat{q_i} = -a_i$. It follows that, in competitive equilibrium, traders simply rid themselves of the hedgeable part of their endowment at zero prices, and end up after the transaction with the part that is independent of the securities. (In this respect, recall the previous Remark 2.1.3.)*

Given that the case $a_I = 0$ is covered by Remark 2.1.4 above, we shall assume tacitly in the sequel that $a_I \neq 0$. (The only point where we return to the case $a_I = 0$ is at Remarks 2.2.1 and 2.3.2.) When $a_I \neq 0$, we define the following parameters, which will turn out to be crucial for our analysis:

$$\beta_i := \frac{\mathbb{C}\mathrm{ov}(E_I, S) C^{-1} \mathbb{C}\mathrm{ov}(E_i, S)}{\mathbb{C}\mathrm{ov}(E_I, S) C^{-1} \mathbb{C}\mathrm{ov}(E_I, S)} = \frac{\langle a_I, C a_i \rangle}{\langle a_I, C a_I \rangle}, \quad i \in I. \tag{2.5}$$

Note that

$$\beta_I \equiv \sum_{i \in I} \beta_i = 1.$$

When the traders' endowments are tradeable, i.e., when the endowment vector E_i belongs in the linear span of $(S_k)_{k\in K}$ for all $i \in I$, then β_i literally coincides with the beta of the ith endowment, in the terminology of the Capital Asset Pricing Model. In general, β_i should be considered as a "*projected beta*" of the ith endowment onto the space of tradeable securities; as stated in the introduction, it shall be called simply (pre-transaction) *beta* in the sequel. Consistent with classical theory, betas shall measure the level of *exposure to market risk* of each trader before and after the equilibrium transaction.

Both equilibrium prices and allocations strongly depend on the traders' heterogeneity. After the competitive transaction, the position of trader $i \in I$ is $E_i + \langle \widehat{q}_i, S - \widehat{p}\rangle$, and one may immediately calculate the post-transaction beta of the position to be equal to λ_i. Hence, at competitive risk sharing, each trader ends up with a positive exposure to market risk, with a beta less than one, even if initial positions are negatively correlated to market risk. Note also that traders with higher risk tolerance are willing to get relatively more exposure to the market risk through the competitive transaction.

The cash amount (*signed risk premium*) that trader $i \in I$ pays to obtain post-transaction beta equal to λ_i is

$$\langle \widehat{q}_i, \widehat{p}\rangle = (\beta_i - \lambda_i)\langle a_I, Ca_I\rangle / \delta_I,$$

which is linearly increasing with respect to β_i. In fact, traders that reduce their beta after the competitive transaction (i.e., those with $\lambda_i < \beta_i$) pay a positive risk premium $\left|\langle \widehat{q}_i, \widehat{p}\rangle\right| = \langle \widehat{q}_i, \widehat{p}\rangle$ to their counter-parties. On the other hand, traders that undertake market risk at the competitive transaction (i.e., those with $\beta_i < \lambda_i$) are compensated with a risk premium $\left|\langle \widehat{q}_i, \widehat{p}\rangle\right| = -\langle \widehat{q}_i, \widehat{p}\rangle$.

Based on the formulas of equilibrium prices and allocations of (2.3) and (2.4), we readily calculate and decompose the traders' utility at competitive equilibrium as

$$\begin{aligned}\mathbb{U}_i\left(E_i + \langle \widehat{q}_i, S - \widehat{p}\rangle\right) &= u_i + \frac{1}{2\delta_i}\left|C^{1/2}(\lambda_i a_I - a_i)\right|^2 = u_i + \frac{1}{2\delta_i}\left|C^{1/2}\widehat{q}_i\right|^2 \\ &= u_i + \underbrace{\frac{1}{2\delta_i}\langle a_i, Ca_i\rangle - \lambda_i^2\frac{\langle a_I, Ca_I\rangle}{2\delta_i}}_{\text{profit/loss from random payoff}} \underbrace{-\frac{\beta_i - \lambda_i}{\delta_I}\langle a_I, Ca_I\rangle}_{\text{(signed) risk premium}}, \quad i \in I.\end{aligned} \tag{2.6}$$

Larger trades at competitive equilibrium result in higher utility gain after the transaction. The above decomposition of utility into risk-sharing gain and risk premium allows one to further analyse the exact sources of utility for each trader, and will prove especially useful later on, when comparing competitive and noncompetitive equilibria.

2.2 Traders' Best Response Problem

2.2.1 The setting of trader's response problem

While it is rather reasonable to assume that pre-transaction betas are publicly known, it is problematic to impose a similar informational assumption on traders' risk profiles. We view risk tolerance as a subjective parameter, and more realistically consider it as *private information* of each individual trader. In the CARA-normal market setting treated here, each trader's risk tolerance is reflected in the elasticity of the submitted demand function. In particular, from Proposition 2.1.2 and the induced individual utility gain (2.6), elasticities of traders' submitted demand directly affect both the allocation of market risk and the associated risk premia. Therefore, it is reasonable to inquire whether an individual trader has motive to strategically choose the elasticity of the submitted demand function. More precisely, adapting the family of linear demand functions with downward slope of the form (2.2), strategically chosen elasticity is equivalent to submitting demand function

$$Q_i^{\theta_i}(p) = -a_i - \theta_i C^{-1} p, \quad p \in \mathbb{R}^K, \tag{2.7}$$

where $\theta_i \in (0, \infty)$ is the elasticity of the submitted demand function $Q_i^{\theta_i}$; equivalently, $1/\theta_i$ is the risk aversion reflected by the submitted demand. In the extreme case where $\theta_i \to \infty$, trader $i \in I$ submits extremely elastic demand, or equivalently behaves as risk neutral, while $\theta_i \to 0$ indicates extremely inelastic demand, i.e., a case where the trader does not want to undertake any risk.

The question addressed in the present section is how traders choose the elasticity of their demand function within the family of demands (2.7), and whether this is different than their risk tolerance. In order to make headway with examining the best response function of trader $i \in I$, we assume that all traders except trader $i \in I$ have submitted an aggregate linear demand function of the form (2.7), where $\theta_{-i} = \sum_{j \in I \setminus \{i\}} \theta_j \in (0, \infty)$ is the aggregate elasticity of all traders except trader $i \in I$. Under this scenario, if trader $i \in I$ chooses to submit the demand function (2.7) with $\theta_i \in (0, \infty)$, and recalling (2.3) and (2.4), the equilibrium price and allocations will equal

$$\widehat{p}(\theta_i; \theta_{-i}) = -\frac{1}{\theta_i + \theta_{-i}} C a_I, \qquad \widehat{q}_i(\theta_i; \theta_{-i}) = \frac{\theta_i}{\theta_i + \theta_{-i}} a_I - a_i,$$

and hence the trader's payoff will equal

$$E_i + \left\langle \widehat{q}_i(\theta_i; \theta_{-i}), S - \widehat{p}(\theta_i; \theta_{-i}) \right\rangle .$$

Since $\theta_{-i} > 0$, the limiting cases when $\theta_i = 0$ (interpreted as extreme inelasticity) and $\theta_i = \infty$ (interpreted as risk neutrality) are well defined; indeed, taking limits in the expressions above, it follows that

$$\widehat{p}(0;\theta_{-i}) = -\frac{1}{\theta_{-i}}Ca_I, \qquad \widehat{q}_i(0;\theta_{-i}) = -a_i,$$
$$\widehat{p}(\infty;\theta_{-i}) = 0, \qquad \widehat{q}_i(\infty;\theta_{-i}) = a_I - a_i = a_{-i}.$$

Risk-neutral acting traders satisfy all the demand of the other traders, accepting all their market risk, without asking a risk premium (recall that we have assumed that $\mathbb{E}[S_k] = 0$, $\forall k \in K$). On the other hand, extremely inelastic demand implies hedging all the initial positions, making the post-transaction beta equal to zero and in fact delegating determination of equilibrium prices to other traders. Using the standard terminology of portfolio management, we call *market-neutral* a position with zero beta.

For $\theta_{-i} \in (0,\infty)$, and under the standing assumption of Gaussian endowments and securities made in Section 2.1, the *response function* of trader $i \in I$ is

$$\begin{aligned}(0,\infty) \ni \theta_i \mapsto \mathbb{V}_i(\theta_i;\theta_{-i}) &\equiv \mathbb{U}_i(E_i + \langle \widehat{q}_i(\theta_i;\theta_{-i}), S - \widehat{p}(\theta_i;\theta_{-i})\rangle) \\ &u_i + \left\langle \frac{\theta_i}{\theta_i+\theta_{-i}}a_I - a_i,\ C\left(\frac{1}{\theta_i+\theta_{-i}}a_I - \frac{1}{2\delta_i}\left(\frac{\theta_i}{\theta_i+\theta_{-i}}a_I + a_i\right)\right)\right\rangle,\end{aligned}$$

with θ_i indicating parametrisation of the trader's strategic behaviour. Since the limiting cases for θ_i are also well defined, we allow a trader to submit demand functions that declare extreme and zero elasticity; for these cases, we have

$$\mathbb{V}_i(0;\theta_{-i}) = \mathbb{U}_i\left(E_i - \langle a_i, S\rangle - \frac{1}{\theta_{-i}}\langle a_i, Ca_I\rangle\right) = u_i + \frac{1}{2\delta_i}\langle a_i, Ca_i\rangle - \frac{1}{\theta_{-i}}\langle a_i, Ca_I\rangle,$$
$$\mathbb{V}_i(\infty;\theta_{-i}) = \mathbb{U}_i(E_i + \langle a_{-i}, S\rangle) = u_i - \frac{1}{2\delta_i}\langle a_{-i}, C(a_I + a_i)\rangle.$$

Summing up, given $\theta_{-i} \in (0,\infty)$, trader $i \in I$'s best response problem is maximising the post-transaction utility by strategically choosing the submitted demand elasticity, i.e.,

$$\theta_i^{\mathrm{r}}(\theta_{-i}) = \underset{\theta_i\in[0,\infty]}{\operatorname{argmax}}\,\mathbb{V}_i(\theta_i;\theta_{-i}). \tag{2.8}$$

Remark 2.2.1 *When $a_I = 0$, $\mathbb{V}_i(\theta_i;\theta_{-i}) = u_i + \langle a_i, Ca_i\rangle/2\delta_i$ holds for all $\theta_i \in [0,\infty]$. In this case, the response function is flat, and any response leads to the same equilibrium prices $\widehat{p}(\theta_i;\theta_{-i}) = 0$ and allocation $\widehat{q}_i(\theta_i;\theta_{-i}) = -a_i$ for trader $i \in I$, irrespectively of the value of θ_{-i}. These are exactly the prices and allocations one obtains at competitive equilibrium.*

The following result shows that, under the assumptions made in Section 2.1 (in particular, that $a_I \neq 0$), the best response problem (2.8) admits a unique solution (recall that β_{-i} denotes the difference $1-\beta_i$, which is equal to $\sum_{j\in I\setminus\{i\}}\beta_j$).

Proposition 2.2.2 *Given $\theta_{-i}\in(0,\infty)$, the best response of trader $i\in I$ exists, is unique and is given as follows:*

$$\theta_i^{\mathrm{r}}(\theta_{-i}) = \begin{cases} 0, & \textit{if } \beta_i \leq -1; \\ \delta_i\theta_{-i}(1+\beta_i)/\left(\theta_{-i}+\delta_i\beta_{-i}\right), & \textit{if } -1<\beta_i<1+\theta_{-i}/\delta_i; \\ \infty, & \textit{if } \beta_i \geq 1+\theta_{-i}/\delta_i. \end{cases} \tag{2.9}$$

Proof: Fix $\theta_{-i}\in(0,\infty)$. Making the monotone change of variable

$$[0,\infty]\ni\theta_i\mapsto k_i := \frac{\theta_i}{\theta_i+\theta_{-i}} \in [0,1],$$

and using a slight abuse of notation, maximising value function $\mathbb{V}_i$ is equivalent to maximising

$$\begin{aligned}\mathbb{V}_i(k_i;\theta_{-i}) &= u_i + \langle a_I, Ca_I\rangle\left(\frac{(1-k_i)k_i}{\theta_{-i}} - \frac{k_i^2}{2\delta_i}\right) - \langle a_I, Ca_i\rangle\,\frac{1-k_i}{\theta_{-i}} \\ &= u_i + \langle a_I, Ca_I\rangle\left(\frac{(1-k_i)k_i}{\theta_{-i}} - \frac{k_i^2}{2\delta_i} - \beta_i\frac{1-k_i}{\theta_{-i}}\right).\end{aligned} \tag{2.10}$$

Since $a_I\neq 0$, the above is a strictly concave quadratic function of $k_i\in[0,1]$; in particular, it has a unique maximum. When $\beta_i\leq -1$ (resp., when $\beta_i\geq 1+\theta_{-i}/\delta_i$), it is straightforward to see that $[0,1]\ni k_i\mapsto\mathbb{V}_i(k_i;\theta_{-i})$ is decreasing (resp., increasing). It follows that $\theta_i^{\mathrm{r}}(\theta_{-i})=0$ when $\beta_i\leq -1$, while $\theta_i^{\mathrm{r}}(\theta_{-i})=\infty$ when $\beta_i\geq 1+\theta_{-i}/\delta_i$. When $-1<\beta_i<1+\theta_{-i}/\delta_i$, first-order conditions in (2.10) give that the unique maximiser of $[0,1]\ni k_i\mapsto\mathbb{V}_i(k_i;\theta_{-i})$ is

$$k_i^{\mathrm{r}}(\theta_{-i}) = \left(2+\frac{\theta_{-i}}{\delta_i}\right)^{-1}(1+\beta_i)\,. \tag{2.11}$$

It then readily follows from (2.11) that the unique maximiser of $[0,\infty]\ni\theta_i\mapsto\mathbb{V}_i(\theta_i,\theta_{-i})$ is $\theta_i^{\mathrm{r}}(\theta_{-i})=\delta_i\theta_{-i}(1+\beta_i)/(\theta_{-i}+\delta_i(1-\beta_i))\in(0,\infty)$. □

According to Proposition 2.2.2, extreme best responses θ_i for trader $i\in I$ are possible, given $\theta_{-i}\in(0,\infty)$. In fact, the best response is zero if and only if $\beta_i\leq -1$, irrespective of the value of θ_{-i}, and the best response is infinity if and only if $\beta_i\geq 1+\theta_{-i}/\delta_i$. In view of this potentiality, it makes sense to understand how a trader would respond if θ_{-i} itself took an extreme value.

We start with the case $\theta_{-i} = \infty$. In this case, taking the limit as $\theta_{-i} \to \infty$ in (2.10) gives

$$\theta_i^{\mathrm{r}}(\infty) = \delta_i(1 + \beta_i)_+. \tag{2.12}$$

The case $\theta_{-i} = 0$ may be treated similarly, but it is worthwhile making an observation. Note that $\theta_{-i} = 0$ means that all other traders except $i \in I$ submit extremely inelastic demands. According to the solution of the best response problem, and anticipating the definition of Bayesian Nash equilibrium in Section 2.3, this only makes sense when $\beta_j \leq -1$ holds for $j \in I \setminus \{i\}$. Since $\beta_i = 1 - \sum_{j \in I \setminus \{i\}} \beta_j$ and there are at least two traders, it should be that $\beta_i > 1$. In this case, taking the limit as $\theta_{-i} \to 0$ in (2.10) gives $\theta_i^{\mathrm{r}}(0) = \infty$. To recapitulate: when $\theta_{-i} = \infty$ the best response is given by (2.12). The case $\theta_{-i} = 0$ is interesting only in the case $\beta_i > 1$, where we set

$$\theta_i^{\mathrm{r}}(0) = \infty, \quad \text{whenever} \quad \beta_i > 1.$$

It is clear from Proposition 2.2.2 that non-price-taking traders have motive to submit demand function of different elasticity than their risk tolerance. The main determinant of departure from the agents' true demand is their pre-transaction beta, defined in (2.5). In order to analyse the effect of strategic behaviour on the equilibrium prices and allocations, we may consider the situation where trader $i \in I$ is the only one acting strategically against price-takers; all other agents submit the elasticity corresponding to their true demand functions for the transaction. In symbols, we set $\theta_{-i} = \delta_{-i}$. This can be seen as a one-sided noncompetitive equilibrium, in the sense that only trader $i \in I$ exploits knowledge on other traders' elasticity and endowments, and responds optimally. The post-transaction beta (2.11) becomes $k_i^{\mathrm{r}} = 0$ when $\beta_i \leq -1$, $k_i^{\mathrm{r}} = 1$ when $\beta_i \geq 1/\lambda_i$, and $k_i^{\mathrm{r}} = \lambda_i(1 + \beta_i)/(1 + \lambda_i)$ when $\beta_i \in (-1, 1/\lambda_i)$. In obvious terminology, we shall call the latter regime *non-extreme*, while the former two will be called *extreme*.

It is completely straightforward from the closed-form expressions for k_i^{r} that

$$\lambda_i < \beta_i \quad \text{if and only if} \quad \lambda_i < k_i^{\mathrm{r}} < \beta_i.$$

Taking into account the discussion following Proposition 2.1.2, the above fact implies that traders have motive to submit more elastic demand functions if and only if they reduce their market risk through the transaction. At the non-extreme regime, this happens when $\beta_i \in (\lambda_i, 1/\lambda_i)$, where the trader's initial position is considered *relatively more exposed to market risk*.

A direct outcome when acting more *aggressively* by submitting more elastic demand is that the post-transaction beta entails more risk: indeed, instead of $\lambda_i \langle a_I, S \rangle$ at competitive equilibrium, the (random part of) the portfolio after submitting demand with elasticity θ_i^{r} equals $k_i^{\mathrm{r}} \langle a_I, S \rangle$. In particular, the post-transaction beta of trader $i \in I$ is k_i^{r}, instead of λ_i. Although the reduction of risk exposure is lower when compared to the competitive equilibrium, it comes at a better price. To wit, we readily calculate that in the whole non-extreme regime $\beta_i \in (-1, 1/\lambda_i)$ it holds that[2]

$$\langle q_i^{\mathrm{r}}, p^{\mathrm{r}} \rangle < \langle q_i^{\mathrm{r}}, \widehat{p} \rangle ,$$

which means that the gain of the strategic behaviour comes from the lower premium that is paid.

Remark 2.2.3 *Under the very special case $\beta_i = \lambda_i$, one obtains $\theta_i^{\mathrm{r}}(\theta_{-i}) = \delta_i$, i.e., $k_i^{\mathrm{r}} = \lambda_i$. In view of* (2.4)*, the latter condition implies $\widehat{q_i} = 0$ and hence trader $i \in I$ does not participate in the sharing of risk; this is also the case in competitive equilibrium.*

2.3 Noncompetitive Risk-Sharing Equilibrium

2.3.1 Nash equilibrium

With the best response problem (2.8) in mind, and assuming that all traders act strategically, we now address noncompetitive Bayesian Nash equilibrium. More precisely, in a fashion similar to the demand-submission game of Kyle (1989), traders submit linear demand schedules of the form (2.7), where $(\theta_i)_{i \in I} \in [0, \infty]^I$ and $\theta_I = \sum_{i \in I} \theta_i > 0$ are the corresponding individual and aggregate submitted demand elasticity. The market equilibrates at the pairs of prices and allocations at which the submitted demands sum up to zero. According to Proposition 2.1.2, as well as relations (2.3) and (2.4), for every submitted demands with elasticities $(\theta_i)_{i \in I} \in [0, \infty]^I$, the prices and allocations that clear out the market are given by $\widehat{p}((\theta_i)_{i \in I}) = -(1/\theta_I) C a_I$, as well as $\widehat{q}_j((\theta_i)_{i \in I}) = (\theta_j/\theta_I) a_I - a_j$, for each $j \in I$. In other words, traders' strategies are parametrised by their submitted elasticity within the family of linear demands (2.7), according to the best response (2.2.2), and noncompetitive equilibria are fixed points of these responses.

[2] When $\beta \in (-1, 1/\lambda_i)$, the exact cash benefit from the best response strategy equals

$$\langle q_i^{\mathrm{r}}, \widehat{p} - p^{\mathrm{r}} \rangle = \langle a_I, C a_I \rangle \frac{\lambda_i (\beta_i - \lambda_i)^2}{\delta_I (1 + \lambda_i)^2 (1 - \lambda_i)}.$$

At competitive equilibrium trader $i \in I$ pays $\langle \widehat{q_i}, \widehat{p} \rangle = (\beta_i - \lambda_i) \langle a_I, C a_I \rangle / \delta_I$ to reduce beta exposure to λ_i, while acting strategically the trader pays $\left\langle q_i^{\mathrm{r}}, p^{\mathrm{r}} \right\rangle = (\beta_i - \lambda_i) \langle a_I, C a_I \rangle (1 - \lambda_i \beta_i)/[(\delta_I - \delta_i)(1 + \lambda_i)^2]$ to reduce beta exposure to k_i^{r}. Note that $\left\langle q_i^{\mathrm{r}}, p^{\mathrm{r}} \right\rangle < \langle \widehat{q_i}, \widehat{p} \rangle$, when $\beta_i \in (\lambda_i, 1/\lambda_i)$.

Definition 2.3.1 *A vector* $(\theta_i^*)_{i\in I} \in [0,\infty]^I$*, with* $\theta_I^* := \sum_{i\in I}\theta_i^* > 0$*, is called* ***Nash equilibrium*** *or* ***noncompetitive equilibrium*** *if, for each* $i \in I$*,*

$$\mathbb{V}_i(\theta_i^*;\theta_{-i}^*) \geq \mathbb{V}_i(\theta_i;\theta_{-i}^*), \quad \forall \theta_i \in [0,\infty].$$

By a slight abuse of terminology, we also call a ***Nash price-allocation equilibrium*** *the corresponding pair* $(p^*,(q_i^*)_{i\in I}) \in \mathbb{R}^K \times \mathbb{R}^{K\times I}$ *given by*

$$p^* = -\frac{1}{\theta_I^*}Ca_I \quad \text{and} \quad q_i^* = \frac{\theta_i^*}{\theta_I^*}a_I - a_i, \quad i \in I. \tag{2.13}$$

where we set $\theta_i^*/\theta_I^* = 1$ *whenever* $\theta_i^* = \infty$*, by convention.*

From the discussion of Section 2.2, and particularly given (2.9) and (2.12), the possibility of noncompetitive equilibrium where some traders behave as being risk neutral (i.e., $\theta_i^* = \infty$ for some $i \in I$) arises. We shall call such Nash equilibria where $\theta_I^* = \infty$ *extreme*, and any other case where the total elasticity θ_I^* belongs to $(0,\infty)$ will be called *non-extreme*.

Remark 2.3.2 *When* $a_I = 0$*, it follows from Remark 2.2.1 that any vector* $(\theta_i)_{i\in I} \in \mathbb{R}_+^I$ *is a Nash equilibrium, always resulting in the same Nash price-allocation with* $p^* = 0$ *and* $q_i^* = -a_i$ *for all* $i \in I$*. Therefore, prices and allocations at competitive and Nash equilibria coincide. In the sequel, we continue the analysis by excluding this trivial case* $a_I = 0$*.*

Remark 2.3.3 *Having defined our notion of noncompetitive equilibrium, we highlight its differences with the thin market models studied in Rostek and Weretka (2015), Malamud and Rostek (2017). As pointed out in the introductory section, the price impact in these papers equals the slope of the aggregate demand submitted by other traders. Traders respond to—or equivalently, trade against—the price impact of their counter-parties forming a slope-game; see (Rostek and Weretka, 2015, Lemma 1) and (Malamud and Rostek, 2017, Proposition 1). Our model keeps the form of equilibrium similar to the competitive one, as the family of demands are linear and of the form* (2.7)*; furthermore, although we parametrise traders' strategies to the single control variable that is elasticity, the key element is that responses, and hence equilibrium conditions, take into account the whole demand function of other traders.*

Our main goal in the sequel is to study existence and uniqueness of the aforementioned linear Bayesian Nash equilibrium, and compare it with the competitive one. Departure from competitive market structure reduces the aggregate transaction utility gain. Indeed, it can be easily checked (see, for example, (Anthropelos and Žitković, 2010, Corollary 5.7)) that the allocation $(\widehat{q}_i)_{i\in I}$ of (2.4) maximises the sum of traders' monetary utilities over all possible market-clearing allocations. As utilities given by (2.1) are monetary, we can measure the risk-sharing inefficiency of any noncompetitive equilibrium as the difference between *aggregate* utility at Nash and competitive equilibrium.

We shall verify in the sequel that risk sharing in the noncompetitive equilibrium is, except in trivial cases, socially inefficient. However, it is not necessarily true that each individual trader's utility is reduced; in fact, it is reasonable to ask which (if any) traders prefer Nash risk sharing in such a thin market, as opposed to the corresponding market that equilibrates in a competitive manner. For this, we compare the individual utility gains at two equilibria, that is, the difference

$$\mathsf{DU}_i \equiv \underbrace{\mathbb{U}_i\left(E_i + \langle q_i^*, S - p^*\rangle\right)}_{\text{utility at Nash equilibrium}} - \underbrace{\mathbb{U}_i\left(E_i + \langle \widehat{q}_i, S - \widehat{p}\rangle\right)}_{\text{utility at competitive equilibrium}}, \quad \text{for each } i \in I, \tag{2.14}$$

and ask when this is positive. Given this notation, and as discussed above, the inefficiency of the noncompetitive risk-sharing is defined as the sum $\sum_{i\in I} \mathsf{DU}_i$.

2.3.2 Equilibrium with at most one trader's beta being greater than one

Under the condition[3] that at most one of the traders have initial beta higher than one, that is

$$\#\{i \in I \mid \beta_i > 1\} \in \{0, 1\}, \tag{2.15}$$

the next result states that there exists a unique *linear* noncompetitive equilibrium.

Theorem 2.3.4 *Under* (2.15)*, there exists a unique Nash equilibrium as in Definition 2.3.1.*

According to (2.9), traders behave as being risk neutral when their initial exposure to market risk is sufficiently higher than one. As we will show in Proposition 2.3.6 below, this behaviour pertains at equilibrium, making it an extreme one, if and only if the following condition holds:

$$\sum_{i\in I} \delta_i (1 + \beta_i)_+ \leq 2 \max_{i\in I}(\delta_i \beta_i). \tag{2.16}$$

[3]We conjecture that Theorem 2.3.4 is true in all cases, although we do not have a rigorous proof of this claim.

When (2.16) fails, the (unique) Nash equilibrium is non-extreme; in this case, and in view of (2.9), the following coupled system of equations

$$\left(2 + \frac{\theta_I^* - \theta_i^*}{\delta_i}\right)\frac{\theta_i^*}{\theta_I^*} = 1 + \beta_i, \quad \forall i \in I \text{ with } \beta_i > -1, \tag{2.17}$$

should hold, where it is θ_I^* which couples the equations. According to (2.9), any trader $i \in I$ with $\beta_i \leq -1$ optimally submits demand function with zero elasticity, inducing a market-neutral post-transaction position, where recall that a position is called market-neutral when it has zero induced beta. Theorem 2.3.4 states, in particular, that the system (2.17) admits a unique solution for an arbitrary number of traders when (2.16) fails. This fact is proved in Appendix A.1, and it is important to note that the proof is *constructive*, and hence can be used to numerically calculate the equilibrium quantities when the number of traders is more than two; the case of two traders admits in fact a closed-form solution and is extensively studied in §2.4.1 later on.

2.3.3 Risk-neutral behaved trader(s)

Having established existence and uniqueness of Nash equilibrium in Theorem 2.3.4, we now show that the condition (2.16) necessarily leads to an extreme noncompetitive equilibrium. We start with an alternative characterisation of(2.16).

Lemma 2.3.5 *Condition* (2.16) *is equivalent to*

$$\beta_k \geq 1 + \frac{1}{\delta_k}\sum_{i \in I\setminus\{k\}} \delta_i (1 + \beta_i)_+, \quad \textit{for some} \quad k \in I. \tag{2.18}$$

Furthermore, (2.18) *can hold for at most one trader* $k \in I$.

Proof: Start by assuming that (2.18) holds, and rewrite it as $\delta_k\beta_k \geq \delta_k + \sum_{i\in I\setminus\{k\}} \delta_i(1+\beta_i)_+$. Since $\beta_k > 1$, which implies that $1 + \beta_k = (1 + \beta_k)_+$, adding $\delta_k\beta_k$ on both sides of the previous inequality and simplifying, we obtain $2\delta_k\beta_k \geq \sum_{i\in I} \delta_i(1+\beta_i)_+$, from which (2.16) follows. Conversely, (2.16) holds if and only if $2\delta_k\beta_k \geq \sum_{i\in I} \delta_i(1+\beta_i)_+$ holds for some $k \in I$. In this case, $\beta_k \geq 0 > -1$, and subtracting $\delta_k(1 + \beta_k) = \delta_k(1 + \beta_k)_+$ we obtain $\delta_k(\beta_k - 1) \geq \sum_{i\in I\setminus\{k\}} \delta_i(1+\beta_i)_+$, which is (2.18).

Assume now that (2.18) held for two traders, say trader $k \in I$ and $l \in I$ with $k \neq l$. Then,

$$\delta_k(\beta_k - 1) \geq \sum_{i\in I\setminus\{k\}} \delta_i(1+\beta_i)_+ \geq \delta_\ell(1+\beta_\ell) \quad \text{and} \quad \delta_l(\beta_l - 1) \geq \sum_{i\in I\setminus\{l\}} \delta_i(1+\beta_i)_+ \geq \delta_k(1+\beta_k).$$

Adding up these inequalities we obtain $-2(\delta_k + \delta_l) \geq 0$, which contradicts the fact that $\delta_k > 0$ and $\delta_l > 0$. We conclude that (2.18) can hold for at most one trader. □

The next result gives a complete characterisation of extreme noncompetitive equilibrium; in particular, it shows that at most one trader—and, in fact, exactly the trader $k \in I$ for which (2.18) holds—may behave as risk-neutral in noncompetitive equilibrium. Note that we do *not* assume (2.15) for Proposition 2.3.6, as it was also not needed for Lemma 2.3.5

Proposition 2.3.6 *An extreme noncompetitive equilibrium (i.e., with $\theta_I^* = \infty$) exists if and only if* (2.16)*, or equivalently* (2.18)*, is true. In this case, we have $\theta_k^* = \infty$ for the unique trader $k \in I$ such that* (2.18) *holds, and $\theta_i^* = \delta_i(1 + \beta_i)_+$ for $i \in I \setminus \{k\}$. In particular, the previous is the unique extreme noncompetitive equilibrium under the validity of* (2.16).

Proof: First, assume that a Nash equilibrium with $\theta_I^* = \infty$ exists. Since $\#I < \infty$, there exists $k \in I$ with $\theta_k^* = \infty$. According to (2.12), for any trader $i \in I \setminus \{k\}$, it holds that $\theta_i^* = \delta_i(1 + \beta_i)_+$. Therefore, for $\theta_k^* = \infty$ to be the best response for trader $k \in I$, (2.9) gives $\beta_k \geq 1 + (1/\delta_k) \sum_{i \in I \setminus \{k\}} \delta_i(1 + \beta_i)_+$. It follows that (2.18) is a necessary condition for existence of an extreme noncompetitive equilibrium.

Conversely, if (2.18) holds, and defining $\theta_k^* = \infty$ and $\theta_i^* = \delta_i(1 + \beta_i)_+$ for $i \in I \setminus \{k\}$, it is immediate from (2.9) and (2.12) to check that the previous is indeed a Nash equilibrium. □

We proceed with some discussion, where we assume that (2.18) holds. In view of Proposition 2.3.6 and the relations in (2.13), at the extreme equilibrium trader $k \in I$ undertakes all market risk, since $q_k^* = a_I - a_k$, and the rest of the traders exchange all their market risk (i.e., $q_i^* = -a_i$, for each $i \in I \setminus \{k\}$) at zero cost, since pricing is done in a risk-neutral way ($p^* = 0$). In particular, the post Nash-transaction beta of trader $k \in I$ reduces to one, and all other traders become market-neutral.

While this transaction is not socially optimal, participating traders increase their utilities; otherwise, equilibrium would not form. Straightforward calculations give the individual utility gains at the extreme equilibrium: $\mathbb{U}_k\left(E_k + \left\langle q_k^*, S - p^*\right\rangle\right) = u_k + \left(\langle a_k, Ca_k\rangle - \langle a_I, Ca_I\rangle\right)/2\delta_k$ and $\mathbb{U}_i\left(E_i + \left\langle q_i^*, S - p^*\right\rangle\right) = u_i + \langle a_i, Ca_i\rangle/2\delta_i$, for each $i \in I \setminus \{k\}$. In particular, the difference of utility gains in (2.14) between the extreme Nash equilibrium and the competitive one equal

$$\mathsf{DU}_k = \frac{\langle a_I, Ca_I\rangle}{2\delta_k}\left[\lambda_k(2\beta_k - \lambda_k) - 1\right], \quad \text{and} \quad \mathsf{DU}_i = \frac{\langle a_I, Ca_I\rangle}{2\delta_i}\lambda_i(2\beta_i - \lambda_i), \ \forall i \in I \setminus \{k\}. \tag{2.19}$$

It follows by straightforward algebra that

$$\text{Risk-sharing inefficiency} := \sum_{i \in I} \mathsf{DU}_i = -\frac{\langle a_I, Ca_I\rangle}{2\delta_I}\frac{1 - \lambda_k}{\lambda_k}.$$

As expected, there is a reduction of the total utility gain when traders behave strategically regarding the elasticity of their submitted demand functions. However, utility gains may be higher in the noncompetitive equilibrium for *individual* traders. From (2.19), we conclude that, in extreme noncompetitive equilibrium, traders that benefit from the market's thinness are the ones with sufficiently high initial exposure to market risk: for trader $k \in I$, when $\beta_k > (1 + \lambda_k^2)/2\lambda_k$ and for traders $i \in I \setminus \{k\}$ when $\beta_i > 2\lambda_i$.[4]

The above quantitative discussion has the following qualitative attributes. Under condition (2.18), in the noncompetitive extreme equilibrium trader $k \in I$ reduces market-risk exposure to one but pays zero premium to other traders. If the market's equilibrium was competitive, trader $k \in I$ would decrease the post-transaction beta even more, to λ_k instead of to one, but the premium would be strictly positive according to the decomposition (2.6). The benefit of zero risk premium prevails the lower reduction of risk if β_k is sufficiently large. On the other hand, the rest of the traders sell all their market-risk exposure at zero premium. For those traders with low initial beta (more precisely, $\beta_i < \lambda_i/2$), the noncompetitive equilibrium leaves them worse off than the competitive one. This stems from the fact that in competitive equilibrium traders with low initial beta obtain premium from traders who are overexposed to market risk, something that does not occur in the noncompetitive extreme equilibrium. However, for traders with $\beta_i \geq \lambda_i/2$, the noncompetitive equilibrium is preferable since they also benefit from the zero risk premium.

To recapitulate: traders that obtain more utility from the extreme noncompetitive equilibrium are the ones with sufficiently high initial exposure to market risk.

2.4 Bilateral Strategic Risk Sharing

2.4.1 The case of essentially two strategic traders

As pointed out in the introductory section, the two-trader case is of special interest since the majority of the OTC transactions consists of only two institutions, or one institution and a client.

Since traders with pre-transaction beta less or equal to -1 always sell all their risk at equilibrium, a risk-sharing game is essentially between two traders if exactly two of them (for concreteness's sake, traders $0 \in I$ and $1 \in I$) have pre-transaction beta larger than -1.

[4]As easy examples show, condition (2.18) does not necessarily imply $\beta_k > (1 + \lambda_k^2)/2\lambda_k$. In the special two-trader case $I = \{0, 1\}$ with $k = 0$, condition (2.18) is equivalent to $\beta_0 > 2 - \lambda_0$, which always implies $\beta_0 > (1 + \lambda_0^2)/2\lambda_0$ when $\lambda_0 > 1/3$. Still in the same two-trader case with $k = 0$, condition (2.18) implies that $\beta_1 < 2\lambda_1$: in the bilateral extreme equilibrium, only the trader that acts as risk neutral could benefit from the market's thinness.

Then, traders 0 and 1 are the only ones to submit demands with non-zero elasticity. In view of the general analysis of §2.3.3, we shall only treat the case of non-extreme equilibrium, i.e., when (2.16) fails. Straightforward algebra yields that, in the present case, failure of (2.16) is equivalent to the following simplified inequality

$$|\lambda_0\beta_0 - \lambda_1\beta_1| < \lambda_0 + \lambda_1. \tag{2.20}$$

If $I = \{0, 1\}$, and recalling that $\beta_0 + \beta_1 = \lambda_0 + \lambda_1 = 1$ in this case, inequality (2.20) is equivalent to $-\lambda_i < \beta_i < 2 - \lambda_i$ for both $i \in \{0, 1\}$.

Proposition 2.4.1 *Assume that $\beta_0 > -1$, $\beta_1 > -1$, $\beta_i \leq -1$ for $i \in I \setminus \{0, 1\}$, and impose* (2.20). *Then there is a unique noncompetitive equilibrium that satisfies $\theta_i^* = 0$ for $i \in I \setminus \{0, 1\}$, as well as*

$$\theta_0^* = \delta_0 \frac{2\lambda_1(\beta_0 + \beta_1)}{(\lambda_0 + \lambda_1) + (\lambda_1\beta_1 - \lambda_0\beta_0)}, \qquad \theta_1^* = \delta_1 \frac{2\lambda_0(\beta_0 + \beta_1)}{(\lambda_0 + \lambda_1) + (\lambda_0\beta_0 - \lambda_1\beta_1)}. \tag{2.21}$$

Proof: As already mentioned, Proposition 2.2.2 implies that the best response for each trader $i \in I$ with $\beta_i \leq -1$ is zero; for traders 0 and 1, θ_0^* and θ_1^* should satisfy (2.17). In this case of essentially two traders, the system takes the form of the following two equations

$$(2\delta_0 + \theta_1^*)\theta_0^* = \delta_0(1 + \beta_0)(\theta_0^* + \theta_1^*) \quad \text{and} \quad (2\delta_1 + \theta_0^*)\theta_1^* = \delta_1(1 + \beta_1)(\theta_0^* + \theta_1^*). \tag{2.22}$$

Subtracting the first equation from the second and dividing by $\theta_I^* = \theta_0^* + \theta_1^*$ gives

$$2(\delta_1 k_1^* - \delta_0 k_0^*) = \delta_1(1 + \beta_1) - \delta_0(1 + \beta_0), \tag{2.23}$$

where $k_i^* \equiv \theta_i^*/(\theta_0^* + \theta_1^*)$ for $i \in \{0, 1\}$. Since $k_1^* = 1 - k_0^*$, (2.23) is a simple linear equation of k_0^* whose unique solution is

$$k_0^* = \frac{1}{2} + \frac{\lambda_0\beta_0 - \lambda_1\beta_1}{2(\lambda_0 + \lambda_1)}. \tag{2.24}$$

The first equation in (2.22) can be written as $(2\delta_0 + \theta_1^*)k_0^* = (1 + \beta_0)\delta_0$, which together with (2.24) implies that θ_1^* should be given as in (2.21). A symmetric argument shows that θ_0^* should also be given as in (2.21). Finally, note that assumption (2.20) and the imposed condition $\beta_i \leq -1$, for each $i \in I \setminus \{0, 1\}$ guarantee that both θ_0^* and θ_1^* are strictly positive and finite. □

At the above noncompetitive equilibrium, prices are given by $p^* = -Ca_I/(\theta_0^* + \theta_1^*)$, while the allocation is $q_i^* = a_I\theta_i^*/(\theta_0^* + \theta_1^*) - a_i$ for each $i \in I$, i.e. only trader 0 and 1 are left with market risk after the transaction.

Remark 2.4.2 *As can be readily checked, a combination of Proposition 2.3.6, Theorem 2.3.4 and Proposition 2.4.1 completely covers all possible configurations for trades including up to three players. On the other hand, one may find a configuration of four traders that is not covered by the results; for example, with* $I = \{0, 1, 2, 3\}$ *and* $\delta_i = 1$ *for all* $i \in I$, *let* $\beta_0 = \beta_1 = 2$, $\beta_2 = 0$, $\beta_3 = -3$.

For the rest of this section we focus our analysis and discussion on bilateral transactions, where we assume that $I = \{0, 1\}$. For the reader's convenience, we note the following result stemming immediately from Proposition 2.4.1.

Corollary 2.4.3 *When* $I = \{0, 1\}$ *and under inequality* (2.20), *there is a unique linear noncompetitive equilibrium given by*

$$(\theta_0^*, \theta_1^*) = \left(\delta_0 \frac{2\lambda_1}{\lambda_1 + \beta_1}, \delta_1 \frac{2\lambda_0}{\lambda_0 + \beta_0}\right).$$

The corresponding price-allocation equilibrium is given by

$$p^* = -\frac{\delta_I(\lambda_0 + \beta_0)(\lambda_1 + \beta_1)}{4\delta_0\delta_1} C a_I = \frac{(\lambda_0 + \beta_0)(\lambda_1 + \beta_1)}{4\lambda_0\lambda_1} \widehat{p}, \tag{2.25}$$

and

$$q_i^* = \frac{\lambda_i + \beta_i}{2} a_I - a_i = \frac{\widehat{q_i}}{2} + \frac{\beta_i a_I - a_i}{2}, \quad i \in \{0, 1\}.$$

Remark 2.4.4 *The only case where the allocation at noncompetitive equilibrium coincides with the competitive one is when* $\beta_0 = \lambda_0$, *which necessarily implies that* $\beta_1 = \lambda_1$ *also holds. This equality, however, means that the competitive equilibrium is a trivial no-transaction equilibrium, since* (2.4) *gives* $q_0^* = 0 = q_1^*$.

As expected from the analysis of Section 2.2, relatively higher initial exposure to market risk implies more higher submitted elasticity at the noncompetitive equilibrium: for each $i \in \{0, 1\}$,

$$\delta_i < \theta_i^* \quad \Leftrightarrow \quad \lambda_i < \beta_i \quad \Leftrightarrow \quad \lambda_{-i} > \beta_{-i}.$$

In particular, the trader who reduces (resp., increases) exposure to market risk through the transaction submits a demand function with higher (resp., less) elasticity than the one that corresponds to that trader's risk tolerance.

The above analysis implies that the trader with higher initial exposure to market risk is willing to retain some of this risk in exchange for a lower risk premium. Correspondingly, the trader who undertakes further market risk through the transaction tends to behave in a more risk averse way, hesitating to undertake more risk at the same risk premium. The direct outcome is that the volume of risk sharing is lower than the one obtained at competitive equilibrium, which leads to *risk-sharing inefficiency*. In fact, simple calculations yield that the Nash post-transaction beta of trader $i \in I$ changes from β_i to $(\lambda_i + \beta_i)/2$, instead of a competitive—and socially optimal—post-transaction beta of λ_i. In other words, for both traders the noncompetitive equilibrium transaction makes their betas exactly equal to the middle point between the initial and the socially optimal ones.

Remark 2.4.5 *From* (2.25)*, we can easily see that* $p^* = \widehat{p}$ *holds if and only if* $\lambda_0 = \beta_0$ *or* $\lambda_0 = (2-\beta_0)/3$. *While the former case is the trivial one (with zero volume at any equilibrium), the latter gives a special non-trivial case where prices remain unaffected by the traders' strategic behaviour. In this case, the Nash post-transaction beta is* $(\lambda_i+\beta_i)/2 = (1+\beta_i)/3 = \lambda_{-i}$ *for both* $i \in \{0, 1\}$.

Similar to the decomposition of utility gains at competitive equilibrium in (2.6), we decompose the corresponding utility gains at noncompetitive equilibrium for $i \in \{0, 1\}$ as

$$\mathbb{U}_i\left(E_i + \langle q_i^*, S - p^* \rangle\right) = u_i + \underbrace{\frac{1}{2\delta_i}\langle a_i, Ca_i\rangle - \left(\frac{\lambda_i + \beta_i}{2}\right)^2 \frac{\langle a_I, Ca_I\rangle}{2\delta_i}}_{\text{profit/loss from random payoff}} \underbrace{-\frac{\beta_i - \lambda_i}{\delta_I}\langle a_I, Ca_I\rangle L}_{\text{(signed) risk premium}}, \tag{2.26}$$

where $L \equiv (\beta_0 + \lambda_0)(\beta_1 + \lambda_1)/8\lambda_0\lambda_1$. The decompositions (2.6) and (2.26) give an expression for the utility difference between the two equilibria DU_i defined in (2.14); to wit,

$$\mathsf{DU}_i = \frac{\langle a_I, Ca_I\rangle}{2\delta_i}\left[\lambda_i^2 - \left(\frac{\lambda_i + \beta_i}{2}\right)^2\right] + \frac{\beta_i - \lambda_i}{\delta_I}\langle a_I, Ca_I\rangle (1 - L), \quad i \in \{0, 1\}\,. \tag{2.27}$$

As was the case in extreme equilibrium discussed in §2.3.3, the difference of utility gains stems from two sources: the gain from sharing the random (risky) payoffs and the risk premium paid or received. Lets assume without loss of generality that $\beta_0 < \lambda_0$ (or equivalently, that $\beta_1 > \lambda_1$), i.e., that trader 0 undertakes more market risk after the (competitive or not) transaction. Since noncompetitive risk-sharing beta reaches only halfway compared to competitive risk-sharing, there is less risk undertaken by trader 0. The risk premium received for undertaking market risk is higher than the one in competitive equilibrium if and only if $L > 1$, which holds in particular when λ_0 is close to one. When λ_0 is not close to one, the risk premium is lower and could absorb all the gain from the lower undertaken market risk. Hence, for traders who undertake market risk at the transaction and have risk preferences close to risk neutrality, the noncompetitive equilibrium is more beneficial.

On the other hand, trader 1 is selling market risk, with lower reduction of Nash post-transaction beta (a fact that decreases utility), while the premium is lower at Nash equilibrium if and only if $L < 1$. The difference $1 - L$ is negative for λ_1 close to zero, and the total difference (2.27) for $i = 1$ remains negative when $\beta_1 < 1$ for every value of λ_1. For fixed λ_1, L is decreasing in β_1 (when $\beta_1 > 1 - \lambda_1$) and the total difference (2.27) for $i = 1$ is positive when β_1 is close to its upper bound $2 - \lambda_1$.

Finally, it should be pointed out that when the risk preferences of trader 0 (i.e., the buyer of market risk) are close to risk neutrality (that is, when λ_0 is close to 1), the noncompetitive equilibrium is always better than the competitive one if and only if $|\beta_0| < 1$ or, equivalently, when $0 < \beta_1 < 2$. In particular, (2.27) and the discussion of extreme equilibrium in §2.3.3 imply that

$$\lim_{\delta_0 \to \infty} \mathsf{DU}_0 = \begin{cases} \langle a_I, C a_I \rangle (1 + \beta_0)(1 - \beta_0)^2 / 8\delta_1, & \text{if } \beta_0 \in (-1, 1); \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, within non-extreme Nash equilibrium, traders that obtain more utility in the noncompetitive equilibrium are the ones with risk preferences close to risk neutrality.

Overall, we conclude that *in two-trader transactions, traders that benefit with more utility from the noncompetitive equilibrium are the ones with sufficiently high initial exposure to market risk, and traders with sufficiently high risk tolerance.*

2.4.2 The effect of incompleteness in thin markets

As emphasised above, our model allows the market to be incomplete, in that the tradeable securities do not necessarily belong to the span of the traders' endowments. When traders' endowments are not securitised, risk-sharing through competitive trading of other securities is sub-optimal. The goal of this section is to examine the effect of a market's incompleteness, both on aggregate and individual levels, when the risk-sharing is *noncompetitive*. For this goal, we consider the indicative two-trader game, $I = \{0, 1\}$.

In order to examine the effect of a market's incompleteness we compare two market settings: an incomplete one, and one where $S = E$. To highlight the effect of incompleteness, we assume that besides (lack of) completeness, the rest of the parameters are the same; in particular, risk aversions remain the same, and projected and actual betas are equal. We take into account the individual utility gains (2.6), (2.26) and utility difference (2.27). For quantities pertaining to the complete market we use notation with superscript "o", that is, (q_i^o, p^o) are the noncompetitive equilibrium allocations and price and $\hat{q}_i^o$ the allocation under competitive equilibrium. Straightforward calculations give the following decomposition of utility gains, in terms of gains in competitive equilibrium and the effect of market noncompetitiveness:

$$\text{Utility gain in incomplete setting} = \mathbb{U}_i\left(E_i + \langle q_i^*, S - p^*\rangle\right) - u_i = \underbrace{\frac{1}{2\delta_i}\left|C^{1/2}\widehat{q}_i\right|^2}_{\text{gain in competitive equilibrium}} + \mathsf{DU}_i$$

$$\text{Utility gain in complete setting} = \mathbb{U}_i\left(E_i + \langle q_i^o, E - p^o\rangle\right) - u_i = \underbrace{\frac{1}{2\delta_i}\left|\mathbb{C}\mathsf{ov}^{1/2}(E,E)\hat{q}_i^o\right|^2}_{\text{gain in competitive equilibrium}} + \mathsf{DU}_i^o.$$

Based on the above, we notice the following: The first term represents the gains of the risk-sharing if the markets were competitive. In particular, we have that (see also Proposition 2.7 in Anthropelos (2017))

$$\left|C^{1/2}\widehat{q}_i\right|^2 = \mathbb{C}\mathsf{ov}(S, \lambda_i E_I - E_i)C^{-1}\mathbb{C}\mathsf{ov}(S, \lambda_i E_I - E_i) \leq \mathbb{V}\mathsf{ar}(\lambda_i E_I - E_i) = \left|\mathbb{C}\mathsf{ov}^{1/2}(E,E)q_i^o\right|^2,$$

where equality holds if, and only if, S belongs in the span of E. The above inequality means that, under a competitive market setting, each trader loses utility due to the market's incompleteness.

The effect of the market's incompleteness on the noncompetitive transaction, after accounting for the differences in the competitive environment, is captured by the difference $\mathsf{DU}_i^o - \mathsf{DU}_i$. In view of (2.27), we have

$$\mathsf{DU}_i = \frac{\langle a_I, Ca_I\rangle}{2\delta_i}\left[\lambda_i^2 - \left(\frac{\lambda_i + \beta_i}{2}\right)^2 + 2\lambda_i(\beta_i - \lambda_i)(1 - L)\right], \quad i \in \{0, 1\}. \tag{2.28}$$

Keeping the parameters β_i, λ_i equal for the complete and incomplete market settings, the only difference stems from the term $\langle a_I, Ca_I\rangle$. In the incomplete market setting this term equals $\mathbb{C}\mathsf{ov}(S, E_I)C^{-1}\mathbb{C}\mathsf{ov}(S, E_I)$, while in the complete market setting it equals $\mathbb{V}\mathsf{ar}(E_I)$. Since

$$\mathbb{C}\mathsf{ov}(S, E_I)C^{-1}\mathbb{C}\mathsf{ov}(S, E_I) \leq \mathbb{V}\mathsf{ar}(E_I), \tag{2.29}$$

market incompleteness decreases (resp., increases) the utility gain (resp., loss) that is caused by the market's noncompetitiveness. In other words, traders that benefit from the noncompetitive market setting (i.e., those with high risk tolerance and/or high exposure to market risk), have their utility gains reduced by the fact that endowments are not securitised. More precisely, we have seen that traders with relatively high exposure to market risk behave as risk neutral in order to reduce their exposure to one without paying risk premium. When the market is complete the reduction of the risk is more effective, since the traders sell part of their endowments and not a security that is simply positively correlated with their endowments, as in the incomplete setting. Recall also that the utility gain of the traders with relatively lower risk aversion under noncompetitive setting stems from the lower (resp., higher) risk premium that they pay (resp., receive). From (2.26), we get that the risk premium is always higher in the complete market setting (see also (2.29)) and hence the aforementioned increase (resp., decrease) of risk premium is also higher in the competitive setting.

We may conclude that, although market's incompleteness reduces the aggregate efficiency of risk-sharing, it also reduces the differences of utility gains/losses among traders.

Chapter 3

The Effect of Market Conditions and Career Concerns in the Fund Industry

3.1 Introduction

In recent years, there has been growing concern in the financial markets about the role of various financial intermediaries such as mutual funds and hedge funds, as the proportion of the institutional ownership of equities has sharply increased and the Global Assets under Management are estimated to exceed $100 trillion by 2020[1]. The managers of these funds are competing with each other, but also with alternative investment vehicles such as market index funds or ETFs, to attract new investors. One of the ways in which they differentiate themselves is through their investment strategy. In particular, managers often signal their confidence by choosing strategies that are highly idiosyncratic[2], and more importantly their incentive to pick these strategies fluctuates with the general market conditions.

Our first contribution is to build a model in which a manager's investment decision provides an imperfect signal on her ability to generate idiosyncratic returns. To be more precise, the manager will skew her investment choice towards a strategy with low exposure to the market in order to signal her confidence. A highly skilled manager is more likely to invest in her idiosyncratic project, since this will deliver on average superior returns. The investors cannot observe directly the manager's ability, but because of the above they will associate an idiosyncratic strategy with a competent manager; in turn, this will endow such a strategy with a reputational benefit. This asymmetry of information between the manager and her potential investors is the main driving force behind the results of this paper.

[1] This is according to a research by PWC.

[2] For example, a recent article in Financial Times explains how institutional investors are turning to alternative investments in recent years.

Our second contribution is to demonstrate that the signalling value of investing in a low beta strategy depends on the market conditions. Managers have a dual objective; they want to maximise their contemporaneous returns but also their perceived reputation. The better the market (*bull*) is, the more the managers face a trade-off between these two objectives, and the less the investors penalise managers for choosing a high beta strategy. Consequently, there is an interaction between managers' career concerns and market conditions.

To analyse the above interactions we consider a two period model in which there is a continuum of investors and a single fund manager. Each investor chooses between investing his wealth through the manager, or directly in the market index, and this choice is affected by an investor's specific stochastic preference shock. The manager's utility is a function of the fees she collects, which are an exogenous proportion of her fund's assets under management (AUM) at the end of each period. After the investors have allocated their funds, the manager publicly chooses between a high or low beta investment strategy. We model the manager's ability as the ex ante expected return of her idiosyncratic strategy, which is either high or low. In each of the two periods, and before picking an investment strategy, the manager also receives a private signal on the contemporaneous profitability of her idiosyncratic project. Both her ability and this signal are her private information, and she uses them to form her final estimate of the profitability of her contemporaneous idiosyncratic strategy. As a result, a high type manager is more likely to form a high estimate, but this is not always the case.

To model market conditions, we assume that the manager also receives a signal on the market's contemporaneous return. This signal is eventually revealed to the investors, but only after they have made their own investment choice. In some sense, we allow for them to eventually understand the market conditions under which the manager acted. However, at this point it will be too late for them to use this information to trade on their own[3]. In section 3.3.4, we extend our setting by allowing two managers to coexist in the market, in order to study how the competition is affected by market conditions. We focus mainly on the first period, since in the second the manager's investment choice is not affected by her reputational concerns. In fact, the second period is introduced in order to create those concerns.

For our first result, we analyse a refinement of the perfect Bayesian Equilibrium, which we call *monotonic equilibrium* and we prove that this always exists. The only additional restriction that this refinement is imposing is that the manager's reputation is non-decreasing on her performance. In addition, under mild parametric restrictions we demonstrate that the monotonic equilibrium is unique.

In our second result we demonstrate that investing in an idiosyncratic strategy carries a reputational benefit. This is because, the cut-off of the high manager type is smaller than that

[3]In other words, manager has a superior market-timing ability compared to an investor.

of the low. In other words the high type is more receptive to the idea of adopting a low beta strategy. Intuitively, the manager's choice is affected by two incentives. On the one hand, she wants to increase her reputation, which skews her preferences towards idiosyncratic investments. On the other hand, she cares about the realised return of her strategy, since her fees depend on it. Hence, for a relatively low private signal even a high type may opt to forfeit the reputational benefit, because investing in the market will generate higher returns, and as a result more fees. Therefore, the investment strategy is informative but it does not fully reveal the manager's ability, which is a realistic representation of the fund industry.

Our third and most important result is to show that the reputational benefit of investing in the idiosyncratic project is decreasing in the market conditions. In particular, we prove that the expected sensitivity of reputation to performance is higher in bear markets than in bull markets. This is because investors understand the dual objective of managers and the fact that a manager is more likely to invest in the market when the market conditions are good, and thus update their beliefs less aggressively when this is the case; instead, in bad times any change in a fund's performance is much more likely to be attributed to the ability of the manager.

We use the above results to discuss the competition between funds, in terms of their sizes, and its fluctuation depending on market conditions. We predict that the likelihood of changes in the ranking of the funds, measured by assets under management, is hump shaped on the market return, but is also higher during bear markets than during bull markets, due to the higher informativeness of performance; we also find some empirical evidence supporting this prediction. This is in line with the common perception that the industry only rearranges its interaction with its investors during crises.

Finally, as an extension to our model, we study the case where investors cannot observe the managers' investment decision. In this scenario, we assume that the investors cannot observe if the manager had invested on the market or their idiosyncratic portfolio, and we conclude that, under this assumption, the conditions for the existence of a monotonic equilibrium cannot be satisfied.

Academic research in financial intermediaries has so far mainly focused on establishing various empirical results about their structure, returns, flows, managers' skill and many other characteristics; there have been far fewer theoretical papers. One of the seminal papers about mutual funds is from Berk and Green (2004); they construct a benchmark rational model in which the lack of persistence of outperformance, is not due to lack of superior skill by active managers, but is explained by the competition between funds and reallocation of investors' capital between them.

Our paper aims to contribute to various strands of literature that we outline below. First, it relates to many papers that study how managers' concerns about their reputation affect their investment behaviour. Chen (2015) examines the risk taking behaviour of a manager who privately knows his ability and shows that in this model investing in the risky project always makes a manager's reputation higher, thus leading to overinvestment in such risky projects. Dasgupta and Prat (2008) study the reputational concerns of managers, and show how they may lead to herding and can explain some market anomalies; their focus though is mainly on the asset pricing implications of this behaviour. Similarly, Guerrieri and Kondor (2012) build a general equilibrium model of delegated portfolio management to study the asset pricing implications of career concerns; they find that as investors update their beliefs about managers, these concerns lead to a reputational premium, which can change signs depending on the economic conditions. Moreover, Malliaris and Yan (2015) show that career concerns induce a preference over the skewness of their strategy returns, while Hu et al. (2011) present a model of fund industry in which managers alter their risk-taking behaviour based on their past performance and show that this relationship is U-shaped. Huang et al. (2012) on the other hand, build a theoretical framework to show how investors are rationally learning about the managers' skills, and test their predictions about the fund flow-performance relationship empirically; however, they do not take into account any strategic behaviour by the fund managers.

The paper most relevant to our work is that by Franzoni and Schmalz (2017). In their work, they study the relationship between the fund to performance sensitivity and an aggregate risk factor and they find that this is hump shaped. They also build a theoretical model in which investors update their beliefs about the managers' skills while they also learn about the fund's exposure to the market. The second inference in extreme markets is noisier for two reasons. The first is idiosyncratic risk and the second is that investors who are uncertain about risk loadings cannot perfectly adjust fund returns for the contribution of aggregate risk realisations. As a result it becomes harder for investors to judge the managers and update their beliefs, and this is what drives the documented result. The theory we propose differs from that of Franzoni and Schmalz (2017) because their model describes the fund's loading on aggregate risk (β) as a preset fund specific exposure, whereas our model gives the ability for managers to strategically choose their investment decision. Also we further investigate how this investment decision will affect the managers' decision if it is observable by the investors or not. Moreover the data source considered for their paper is the CPRSP Mutual Fund Database which is different from the Morningstar CISDM which we use for the empirical part, making it difficult to compare our results. Although the implementation and the structure of their model is completely different to ours and does not imply the same predictions we are making, we conclude that the aggregate risk realisations matter for mutual fund investors and managers.

Another strand of literature in which we contribute to, is the empirical research on the fund flows and characteristics. It is well documented that mutual fund investors chase past returns, Ippolito (1992) and Warther (1995) present empirical evidence supporting our predictions. Sirri and Tufano (1998) show that the flow-performance relationship is convex, and asymmetrically so on the positive side of returns. Furthermore, Chevalier and Ellison (1997), show that managers engage in window dressing their portfolios. More recently, Wahal and Wang (2011) study the competition between funds, by looking at the effect of the entry of new mutual funds on fees, flows and equilibrium prices. Finally, Ma (2013) provides a very comprehensive survey of empirical findings concerning the relationship between mutual fund flows and performance.

The rest of the paper is organised as follows. In section 3.2, we introduce our theoretical framework and our equilibrium. Section 3.3 proves its existence, identifies a condition under which this is unique, and presents our theoretical predictions. In particular, section 3.3.4 discusses the implications of adding a second manager. Subsequently, section 3.4 presents our empirical results. Section 3.5 considers an alternative model where the investment decision is unobservable. Finally, section 3.6 concludes.

3.2 The Model

3.2.1 Setup

This is a two period model $t \in \{1, 2\}$. There is one fund manager (she) and a continuum of investors (he) of measure one, who collectively form the market. The manager discounts the future with $\delta \in (0, 1]$.

At the beginning of period t, each investor decides how to invest a unit of wealth. At the end of period t, he consumes all the wealth that this investment generated. The investor is restricted to a binary decision. He can either opt to allocate all his wealth in an index tracking strategy. This has the same returns as the market portfolio, which is given by

$$m_t \sim \mathcal{N}(\mu, \sigma_m^2) \tag{3.1}$$

Or, he can choose to invest all his wealth in the manager's fund[4]. For each unit of wealth invested with the manager let $R_t = \exp(r_t)$ denote its value at the end of this period, where

$$r_t = (1 - \beta_t) \cdot a_t + \beta_t \cdot m_t \tag{3.2}$$

is the fund's return. This has two components, one of which is the market return m_t. The second is given by

$$a_t \sim \mathcal{N}(\alpha, \sigma^2) \tag{3.3}$$

which represents the market neutral component of the manager's investment strategy[5]. Adhering to the fund industry's convention, the manager's ability to create idiosyncratic profits is called alpha, and is represented by $\alpha \in \{L, H\}$ where $L < H$. The manager's ability is her private information. The investors share the public prior $\pi = \mathbb{P}(\alpha = H)$.

Finally, β_t represents the fund's exposure to the market. This is publicly chosen by the manager after the investors have allocated their wealth. For simplicity we assume that $\beta_t \in \{0, 1\}$. Note that the model's beta β_t despite its relevance to the corresponding variable

[4]Our underlining intuition is that most of the market participants follow a rule of thumb to their investment through intermediaries. For example, they set apart 5% of their wealth and then they decide if they should invest this amount to a fund.

[5] For example think of a long/short equity fund that invests $(1 - \beta_t)$ of its assets on a market neutral portfolio and β_t on the S&P 500 index. For the most part we refrain from giving a specific interpretation of the components of the fund's return r_t, or which part of its investment strategy they represent. Our framework relies on the simple intuition that some of the return generated by the manager stems from her own ability and some from factor loading. In fact m_t could represent any such factor, and for some funds other choices would be more sensible. For example, a macro fund is more related to the risk-free interest rate than to the equity markets.

of the CAPM model, is not the same variable. Rather the former represents a deterministic investment decision, whereas the latter its estimate.

In addition, before making her investment decision β_t, but after the investors have allocated their wealth, the manager receives two signals

$$s_t \sim \mathcal{N}(a_t\,,\,v^2\,) \qquad \text{and} \qquad s_t^m \sim \mathcal{N}(m_t, v_m^2). \tag{3.4}$$

On the one hand, s_t is private and it is associated to the manager's contemporaneous confidence on her alpha[6]. On the other hand, s_t^m is public but it only becomes available after the investors have committed their capital to the manager's fund. This market signal is considered to be the standard piece of information that most institutional participants receive on the market's condition.

To simplify matters, we assume that the manager's fees are exogenously set to a given percentage $f_t \in [0, 1]$ of her asset under management (AUM) at the end of t[7]. Even though we do not allow for incentive fees, the plain managerial fees f_t we consider suffice to create direct incentives for the manager to perform in t, as her period income per dollar invested is $f_t R_t$.

Two more important assumptions have been made. First, that the manager's investment decision is binary. In particular, it allows for either investing all of the fund's assets in the manager's idiosyncratic strategy a_t, or all in the market m_t. Second, that this decision is observable by the rest of the market participants. The former assumption is imposed mainly to make the model more tractable. We speculate that altering it to allow for $\beta_t \in \{\underline{b}, \overline{b}\}$, where $\underline{b} < \overline{b}$, would not affect our results qualitatively[8]. Regarding the latter assumption, it appears to be reasonable for long investment horizons. This is because the fund's exposure to the market can be ex-ante approximately inferred, either by estimating a multi-factor regression, or by looking at its past portfolio composition, which in many cases is public.

[6]This could reflect the fact that her idiosyncratic strategy has some seasonality that she is able to partly predict. Another interpretation is that the strategy itself changes across periods, in which case α represents the manager's latent ability to come up with new ideas to beat the market.

[7]Endogenising the choice of fees is left for future research. The complexity of allowing an endogenous choice is that the fees would then serve as a signalling device for the managers' ability, thus making the equilibrium much harder to find.

[8] A possibility that we exclude and is worth mentioning is that of a manager that bets against the market. In particular, in strong bear markets most funds would prefer to short the market portfolio, instead of adopting a strategy that is neutral to it. This would have a significant impact on our analysis. Despite that, it is ignored both to facilitate the exposition and because funds that systematically hold big negative positions are not that common.

3.2.2 Payoffs

Investors are risk-neutral, however each one's decision is influenced by an exogenous preference shock which follows an exponential distribution.

$$z_t^j \sim \exp(\lambda)\,, \quad \text{where} \quad j \in [0,1] \tag{3.5}$$

stands for the shock on investor's j preferences at period t and λ is the rate parameter of the exponential distribution. Hence, his payoff from investing in $i \in \{1, m\}$ is

$$v(i, z_t^j) = \begin{cases} \exp(z_t^j - \bar{z}) \cdot (1 - f_t) \cdot R_t & , i = 1 \\ \exp(m_t) & , i = m \end{cases} \tag{3.6}$$

where $\bar{z} > 0$ is a constant that we introduce to ensure that under the lowest preference shock $z_t^j = 0$ the investor would opt for the market instead.

There is a plethora of ways to interpret this shock, a valid one being that each investor values specific fund characteristics, for example the fund's classification with regards to its investment strategy, its portfolio composition, leverage, etc. An alternative one would be that he is influenced by interpersonal relationships, network effects, word of mouth, or other forms of private information. Our analysis will be silent as to what generates this shock.

Furthermore, note that because R_t comes from a log-normal distribution, we could adopt a CRRA utility function for the investor without altering his decision significantly. However, we opt not to do so in order to maintain our expressions as compact as possible. On the other hand, it will be assumed that the manager has log preferences. In particular, if A_t stands for the AUM the fund in the beginning of t, then manager's payoff at t is $\log(A_t f_t R_t)$. Again we speculate that most of our results would not be significantly different if a generic CRRA was used instead of log, however it turns out that this is the most convenient functional form to work with.

3.2.3 Timing

To sum up, the timing in our model is as follows. In each period $t \in \{1, 2\}$, first the preference shock z_t^j, $j \in [0, 1]$, is realised and then the investors decide how to allocate their wealth. Second, the manager receives the private and public signals s_t and s_t^m, respectively. Third, the investment decision β_t is made by the manager, R_t is realised, and both become public. Fourth, the fund's AUM is divided between the manager and her investors, according to the fee f_t, and is consumed immediately. Finally, we assume that the investors that are active in the second period observe the public variables of the first period before allocating their wealth. Importantly, they know (R_1, β_1, s_1^m) and use them to update their beliefs on the manager's ability α. Signal s_1 can not be used since it is private information of the manager and it will never be known to the investors.

3.2.4 Monotonic equilibrium

We call an equilibrium of our model *perfect Bayesian* (PBE), if all market participants use Bayes' rule to update their beliefs on α, whenever possible, and choose their actions in order to maximise their expected discounted payoff at each point they are taking an action. There is a possibility of there being multiple equilibria, which is a common setback for these types of models. For this reason we will further refine the set of equilibria using the following definition, however, the study of these equilibria is beyond the scope of this paper.

Definition. Call a PBE a *monotonic equilibrium* if the manager's reputation, for a given choice of investment strategy, is non-decreasing on her performance.

In other words a monotonic equilibrium satisfies $\mathbb{P}(\alpha = H \,|\, r, s^m, \beta)$ is increasing in r. Therefore, the only requirement that our refinement imposes is that the manager's reputation is not penalised by the fact that she delivers good returns for her investors. The above definition implies that there exists φ_0 and φ_1 such that the public posterior on the manager's ability is given by

$$\begin{aligned} \varphi_0 &= \mathbb{P}(\alpha = H \,|\, r_1, s_1^m, \beta_1), \quad \text{for } \beta_1 = 0 \\ \varphi_1 &= \mathbb{P}(\alpha = H \,|\, r_1, s_1^m, \beta_1), \quad \text{for } \beta_1 = 1 \end{aligned} \tag{3.7}$$

We separate the posteriors that follow each choice of β_1 because those will turn out to have different functional forms.

3.3 Analysis

We begin our analysis by first discussing the manager's optimal investment strategy in the second period and how this affects her career concerns in the first period. Second, we characterise the monotonic equilibrium and prove its existence and uniqueness. Third, we present our results on the baseline model with the single manager. Fourth, we discuss the implications of adding a second manager.

3.3.1 Investment and AUM in the second period

Here we provide a description of how we solve for the manager's investment decision in the second period and the corresponding AUM that this implies. The interested reader can find a more detailed analysis in Appendix B.2.

In the second period the manager faces no career concerns. Hence, the objective of her investment decision is to maximise the expected fees she collects at the end of this period. Because those fees are proportional to her fund's AUM at the end of the second period, and we have assumed log preferences, the manager's payoff maximisation problem simplifies to

$$\max_{\beta_2 \in \{0,1\}} \mathbb{E}[\log(A_2, f_2, R_2) \,\big|\, \beta_2,\ \alpha,\ s_2,\ s_2^m\,]$$

When opting for her idiosyncratic strategy $\beta_2 = 0$ the above expectation uses the manager's ability α and private signal s_2, whereas the index tracking strategy $\beta_2 = 1$ depends only on the market signal s_2^m. Since we have assumed that the returns and the corresponding signals are log-normally distributed we can calculate the above expectation for each choice in closed form. This suggests that the manager's optimal second period strategy is to invest in her idiosyncratic project if and only if $s_2 \geq c(\alpha, s_2^m)$ where

$$c(\alpha, s_2^m) \;=\; \frac{\psi_m}{\psi} \cdot s_2^m + \frac{1-\psi_m}{\psi} \cdot \mu - \frac{1-\psi}{\psi} \cdot \alpha \tag{3.8}$$

The constants ψ and ψ_m are the weights that the Bayesian updating gives to the signals s_2 and s_2^m, respectively, and their functional form can be found in Appendix B.2. Given the above cut-off strategy we can calculate the expected terminal value of one unit of wealth that is invested by the manager. For a high and low type we will denote those by u_2^H and u_2^L, respectively. Therefore, for given posterior reputation φ, and while ignoring the preference shock z, the expected payoff of an investor that opts for the manager is given by

$$[1 - f_2] \cdot [\varphi \cdot u_2^H \;+\; (1-\varphi)\, u_2^L]$$

This together with the assumed preference shock allows us to calculate the assets of the second period in closed-form.

From (3.6) we have the expected payoff of an investor who chooses to invest in a fund or in the market. He chooses the former if his expected payoff is higher. Since there is a continuum of investors with one unit of wealth, the probability of this event occurring is equal to the assets of fund one. Hence

$$A_2(\varphi) \;=\; \left(e^{-(\mu+\bar{z}+\sigma_m^2/2)} \cdot [1-f_2] \cdot [\varphi \cdot u_2^H \;+\; (1-\varphi) \cdot u_2^L]\right)^{\lambda} \tag{3.9}$$

which is an increasing function of the manager's reputation φ. One thing we can note is that as long as $\lambda > 1$, the assets under management are a convex function of the reputation φ. This is a result that has been widely documented in the relevant empirical literature, in slightly different forms.

3.3.2 Existence and uniqueness of the monotonic equilibrium

In this section we demonstrate that the monotonic equilibrium exists and is unique under mild conditions. First, we want to understand the manager's incentives in the first period. Her expected discounted payoff at this point is

$$\mathbb{E}_R\Big[\log\left[R_1 f_1\, A_1(\pi)\right] \;+\; \delta \cdot \log\Big[R_2 f_2 A_2(\varphi_\beta)\Big] \;\Big|\; s^m,\; s,\; \beta\,,\alpha\,\Big]$$

where the expectation taken with respect to the returns of both periods. $A_1(\pi)$ is the equilibrium allocation of AUM in the first period, which has a functional form similar to that of $A_2(\varphi_\beta)$ and β is β_1.

Hereafter, the focus of the paper shifts to the interactions of the first period. As a result, in order to make our formulas more compact, the time subscript t is dropped, whenever this does not create an ambiguity. Using the properties of the natural logarithm we simplify the manager's payoff maximisation problem in period 1 to

$$\max_{\beta\in\{0,1\}}\; \mathbb{E}_r\Big[r \;+\; \delta \cdot \lambda \cdot [\varphi_\beta(r, s^m) \cdot (u^H - u^L) \;+\; u^L\,] \;\Big|\; s^m,\; s,\; \beta\,,\alpha\,\Big] \tag{3.10}$$

Therefore, the manager cares both about her returns in the first period r, but also on how those affect her posterior reputation $\varphi_\beta(r, s^m)$. This reputation is important because it affects the amount of AUM that the manager will manage to gather in the beginning of the second period.

First, we want to offer a characterisation of the monotonic equilibrium.

Lemma 1 *In any monotonic equilibrium the high and low type invest in their idiosyncratic strategy if and only if*

$$s \;\geq\; h(s^m) \quad \text{and} \quad s \;\geq\; l(s^m)\,, \tag{3.11}$$

respectively. At the cut-off the manager should be indifferent between choosing to invest in her idiosyncratic strategy ($\beta = 0$) or in the market ($\beta = 1$). Therefore, the expected utilities *in the corresponding cases should be equal:*

$$\mathbb{E}_r\Big[a + \delta \cdot \lambda \cdot [\varphi_0(r, s^m) \cdot (u^H - u^L) + u^L] \,\Big|\, s = h(s^m),\ \alpha = H\Big]$$
$$= \mathbb{E}_r\Big[m + \delta \cdot \lambda \cdot [\varphi_0(s^m) \cdot (u^H - u^L) + u^L] \,\Big|\, s^m\Big]$$

This implicitly defines $h(s^m)$*. Similarly:*

$$\mathbb{E}_r\Big[a + \delta \cdot \lambda \cdot [\varphi_0(r, s^m) \cdot (u^H - u^L) + u^L] \,\Big|\, s = l(s^m),\ \alpha = L\Big]$$
$$= \mathbb{E}_r\Big[m + \delta \cdot \lambda \cdot [\varphi_0(s^m) \cdot (u^H - u^L) + u^L] \,\Big|\, s^m\Big]$$

which defines $l(s^m)$*.*

This implies that the corresponding conditional expectations, $E[r|s = h(s^m), \alpha = H]$ *and* $E[r|s = h(s^m), \alpha = L]$*, are equal. That is*

$$(1 - \psi) \cdot H + \psi \cdot h(s^m) = (1 - \psi) \cdot L + \psi \cdot l(s^m)$$

and hence

$$l(s^m) - h(s^m) = \frac{1 - \psi}{\psi} \cdot (H - L) \tag{3.12}$$

Proof: In Appendix B.1. □

Hence the more confident the manager becomes on her alpha, the more likely she is to use her idiosyncratic strategy, instead of the index tracking one. In addition, the fact that the high type's cutoff is lower captures the fact that a competent manager uses her idiosyncratic investment strategy relatively more often.

Second, we want to calculate the manager's posterior reputation after each investment decision as a function of her performance.

Lemma 2 (Posteriors) *In any monotonic equilibrium the manager's posterior reputation in the beginning of the second period, if she invested on her alpha* $\beta_1 = 0$ *in the first, is*

$$\varphi_0(r, s^m) = \left(1 + \frac{1 - \pi}{\pi} \cdot \rho(r) \cdot \frac{\Phi\left(\frac{r - l(s^m)(1+\psi) + L\psi}{\nu\sqrt{1+\psi}}\right)}{\Phi\left(\frac{r - h(s^m)(1+\psi) + H\psi}{\nu\sqrt{1+\psi}}\right)}\right)^{-1}, \tag{3.13}$$

where

$$\rho(r) \;=\; \exp\left(\frac{-2(H-L)\,r + H^2 - L^2}{2\nu^2\psi(1+\psi)}\right).$$

On the other hand, if she invested in the market $\beta = 1$ then this becomes

$$\varphi_1(s^m) \;=\; \left(1 + \frac{1-\pi}{\pi} \cdot \frac{\Phi\left(\frac{l(s^m)-L}{\nu}\right)}{\Phi\left(\frac{h(s^m)-H}{\nu}\right)}\right)^{-1} \tag{3.14}$$

We recall that r depends on the investment decision β.

Proof: In Appendix B.1. □

The investors form their posterior belief on the manager's ability by observing her investment decision β and the realised return r. Note that when using her idiosyncratic investment strategy the manager's performance r is generated by her alpha. Hence, in this case the realisation r carries additional information on the manager's ability. On the other hand, when using the index tracking strategy r is equal to the market's return m, which carries no additional information on the manager's ability. This is why φ_0 is a function of r, but φ_1 is not.

Using the above two lemmas, we prove the main result of this part.

Proposition 1 *A monotonic equilibrium always exists. Moreover, a sufficient condition for it to be unique is that*

$$\delta \cdot \lambda \cdot (H-L) \;\leq\; \psi^2 \cdot \nu^2 \tag{3.15}$$

Proof: In Appendix B.1. □

We believe that (3.15) is satisfied for a wide range of parametric specifications that we would consider natural given the economic setting we study. This translates into two requirements. First, that the difference between the ability of the two types is not too big. Second, that the precision of the signal s is neither so small that it becomes irrelevant, nor so big that the manager's ex-ante ability α becomes irrelevant instead.

3.3.3 Results

Here, we present some important properties of the unique monotonic equilibrium. We assume throughout that (3.15) holds. To maintain the notation as light as possible keep using $\varphi_0(r, s^m)$ and $\varphi_1(s^m)$ to refer to the equilibrium reputations, which are obtained after substituting the corresponding values for $h(s^m)$ and $l(s^m)$.

Proposition 2 (Point-wise dominance) *There is a strict reputational benefit for the manager from investing in her alpha, that is*

$$\varphi_0(r, s^m) \; > \; \varphi_1(s^m), \quad \textit{for all} \quad r, s^m \in \mathbb{R}. \tag{3.16}$$

Proof: In Appendix B.1. □

We already know that in every monotonic equilibrium $\varphi_0(r, s^m)$ is increasing in r, in other words high performance is beneficial for the manager's reputation. The proof demonstrates the result by considering the worst case scenario for the manager $\beta = 0$. In the extreme scenario where return approaches minus infinity her reputation is still greater than choose $\beta = 1$. Hence the equilibrium difference between the cutoffs used by the high and low type is such that the investors' inference on the manager's type relies relatively more on her choice of strategy than on the subsequent performance of her fund.

This may seem counterintuitive at first, but it has a very simple explanation. In the appendix we show that for a monotonic equilibrium to also be rational the difference between the equilibrium cutoffs $l(s^m)$ and $h(s^m)$ cannot be too large. If that was the case, then a low type would have to be so confident in order to invest in her alpha that a very bad performance, under the low beta strategy, would be associated with a high type. An immediate consequence of which would be that the manager's reputation would be non-monotonic on her performance. But those are exactly the type of equilibria that appear to be the less realistic.

The above claim is the most challenging one to verify in the data. This is because for each fund we never observe the counter-factual, that is how the fund's flow would look like if it had chosen a lower, or higher beta strategy. Moreover, the simplifying assumption $\beta \in \{0, 1\}$ makes this result stronger than what an alternative model, where the two betas are closer to each other, would give. Despite that, we can verify empirically that to a certain extent a low beta strategy creates enough signalling value to counter the effect of a low subsequent performance.

As a direct consequence of point-wise dominance, we can now get the following interesting proposition, which characterises the effect of the manager's career concerns on her investment behaviour.

Proposition 3 (Investment Behaviour) *The equilibrium cutoffs $h(s^m)$ and $l(s^m)$ are decreasing in the discount factor δ. Moreover, there is overinvestment in the manager's idiosyncratic project, that is*

$$h(s^m) \; \leq \; c(H, s^m) \qquad \textit{and} \qquad l(s^m) \; \leq \; c(L, s^m). \tag{3.17}$$

Proof: In Appendix B.1. □

The proof is a simple application of the implicit function theorem on equation (B.17), the solution of which is shown in the proof of Proposition 1 to be $h(s^m)$. The corresponding result for $l(s^m)$ is obtained by invoking the fact that in every monotonic equilibrium those two cutoffs are connected through a linear relationship, which was again demonstrated in the above proof.

We use the term *"over-investment"* to describe the fact that the manager invests in her idiosyncratic strategy more often than in the absence of career concerns. In other words, over-investment exists when the manager "lower her standards" with regards to her private signal, i.e. she lowers the confidence level required for her to choose the idiosyncratic investment. Note that the manager's optimal cutoff, in the absence of career concerns, corresponds to that already derived from for the second period in (3.8). This is because it is generated by the inefficiency in the investment decision that the manager's career concerns create, which is connected to the underlying parameter δ.

The above proposition demonstrates that there is a bias towards active management in the financial intermediation industry, which is due to its inherent informational asymmetries. To be more precise, we expect managers to get on average less exposure to the market than what would maximise the fund's expected return. Moreover, this action is associated with competence and it is rewarded with an increase in the fund's AUM. Hence, our model provides a theoretical justification for this well documented fact.

Next, we want to see how this bias depends on the unobserved, to the econometricians, market signal s^m and the manager's prior reputation π.

Proposition 4 *The cutoffs $h(s^m)$ and $l(s^m)$ are increasing in the market signal s^m. In addition, there exist lower bounds $\bar{s}^m$ and $\bar{\pi}$ such that for every (s^m, π) such that $s^m \geq \bar{s}^m$ and $\pi \geq \bar{\pi}$ both cutoffs $h(s^m)$ and $l(s^m)$ are increasing functions of the manager's prior reputation π.*

Proof: The proof of the first statement is similar to that of Proposition 3. The proof of the second follows from Lemma 7, which can be found in Appendix B.1. □

The first statement is a very intuitive result. The better the manager expects the market portfolio to perform, the more eager she becomes to invest in it, which translates into higher equilibrium cutoffs.

The crucial implication of the proposition's second statement is that the bias created from the signalling value, of investing in the idiosyncratic strategy, is decreasing in the manager's prior reputation. This is because the equilibrium cutoffs are bounded above by the expected return maximising cutoff $c(\alpha, s^m)$, hence the more π increases the closer they get to it.

A caveat of this result is that it only holds for a manager that is already relatively recognised in the market, in particular it is shown in the appendix that we need at least

$\pi > 1/2$. Intuitively, the closer the prior is to either zero or one, the less it is affected by the actions of the manager. To make this more concrete, think of the extreme case where $\pi \to 1$, in which case it is very difficult for the investors to change their opinion about her ability, as they already know it with almost total certainty. Hence, there is a corresponding result that can be stated for managers of very low reputation. Even though in our model we allow for funds of small size to stay active, in reality most of them would either shut down, or would not even be reported in most datasets, hence we focus just on funds with reputation greater than a 1/2. [9].

Another interesting feature of the presented specification is that it provides a better understanding on how the sensitivity of the fund's asset flows to its performance depend on the market conditions. Let $\varphi^i(r^i, s^m, \beta^i)$ stand for the manager's reputation in either of the two cases and call $\mathrm{d}\varphi^i/\mathrm{d}r^i$ its *sensitivity* with respect to her performance.

Proposition 5 *The conditional probability that a manager has invested in the market portfolio* $\mathbb{P}(\beta^i_t = 1 \,|\, m_t)$ *is increasing in its contemporaneous performance* m_t.

In addition, for a sufficiently reputable manager the conditional expected sensitivity of the manager's reputation with respect to her performance, i.e. $\mathbb{E}_{s^m}[\mathrm{d}\varphi^i/\mathrm{d}r^i \,|\, m_t]$, *is decreasing in* m_t.

Proof: In Appendix B.1. □

When markets are expected to perform well, the manager's direct incentives outweigh those of career concerns. Hence we know from Proposition 4 that she is more likely to give up the reputational benefit of following a low beta strategy. But high beta strategies carry no information with respect to the manager's ability. Hence, even though as noted in Proposition 2 investing in low beta always has a reputational benefit, this benefit is less pronounced in good markets. Therefore investors are expected to rely more on a manager's performance to update their belief about the ability of the manager, when markets are bear than when markets are bull. This result is also supported by the empirical evidence we provide in section 3.4.

3.3.4 Discussion on the competition between funds

It follows from the previous discussion that managers will be judged much more strictly on their performance in bear markets than in bull markets. This in turn has some implications for the relative ranking of the various funds with respect to their reputation, or equivalently their AUM.

[9] Despite that we hope to test empirically if we can obtain a corresponding result for the flows of small funds.

To study this we extend our model by allowing a second manager to operate in the market. We formally define the investor's preference shock in this case and derive the corresponding AUMs of the two funds in Appendix B.2. In fact, the whole analysis of this paper and all our results remain unchanged with the addition of a second manager. The reason is that the manager's utility is such that it is only a function $\varphi_\beta(r, s^m) \cdot (u^H - u^L) + u^L$ and is independent of the number of managers that exist in the model[10].

Our main aim is to study the likelihood of a change in the rank of managers, in terms of investors' beliefs about their ability and relate that to the market conditions. In what follows, we explain why this effect is not monotonic in m_t[11].

Let $\mathbb{P}(i, j \,|\, s^m) = \mathbb{P}(\beta_{\text{fund 1}} = i, \beta_{\text{fund 2}} = j \,|\, s^m)$. In the Appendix it is shown that:

$$\mathbb{P}(\varphi^1 > \varphi^2 \,|\, s^m) \;=\; \mathbb{P}(\varphi^1_0 > \varphi^2_1 \,|\, s^m)\,\mathbb{P}(0, 1 \,|\, s^m) + \mathbb{P}(\varphi^1_0 > \varphi^2_0 \,|\, s^m)\,\mathbb{P}(0, 0 \,|\, s^m), \tag{3.18}$$

What this equation suggests is that the ranks of managers can change through two possible scenarios. In the first scenario, with probability $\mathbb{P}(0, 1 \,|\, s^m)$, one of the two managers invests in his idiosyncratic portfolio and the other follows the market. This probability approaches zero for both very large and very small s^m, as then both managers invest in the market or both invest in their own project. In turn, this makes the first term of the right hand side of the equation (3.18) hump-shaped in s^m. Under this scenario, manager 1 has a reputational benefit from choosing $\beta = 0$ (see Proposition 2) which then makes it possible for his ex-post reputation to be higher than that of manager 2 (despite his initial disadvantage, in terms of the priors π^1, π^2); clearly the smaller the distance between their prior reputations, $\pi^2 - \pi^1$, the larger this likelihood will be.

In the second scenario, with probability $\mathbb{P}(0, 0 \,|\, s^m)$ both managers invest in their own project and manager one receives a much higher return than the other, thus overcoming the effect of the initial prior reputations; in other words, since $\pi^1 < \pi^2$, in order for the posterior reputations to have the opposite order, what needs to happen is that the realised return of manager 1 is much higher than that of 2. This is clearly not possible if they both invest in the market. However, when they both invest in their idiosyncratic project this can happen either because one is luckier than the other, or simply because manager one has high skill and manager two has low skill. This scenario is less likely to occur as the market conditions get better since $\mathbb{P}(0, 0 \,|\, s^m)$ is decreasing in s^m, as we can see from Proposition 5. Moreover, we can get the following remark:

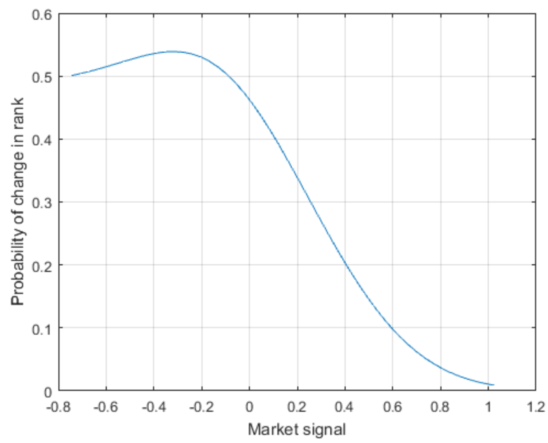
[10]In particular, equation (39) and thus the determination of the cutoffs l and h will remain the same.
[11]Note, we always condition on s^m as we know that all investors observe this market signal.

Remark 1 *The likelihood of a change in the ranks of managers is higher in a very bad market, than in a very good market. That is:*

$$\lim_{s^m \to -\infty} \mathbb{P}(\varphi^1 > \varphi^2 \mid s^m) > \lim_{s^m \to +\infty} \mathbb{P}(\varphi^1 > \varphi^2 \mid s^m) \tag{3.19}$$

The proof of this remark is quite simple. As the market becomes really good, the probability of a manager investing in his own project goes to zero, and hence from (4.20) we see that the probability of a rank change will tend to zero. In contrast, for a very negative market signal, this probability is strictly negative, since $\mathbb{P}(0, 0|s^m) = 1$ and $\mathbb{P}(\varphi_0^1 > \varphi_0^2 \mid s^m) > 0$[12];

From the above analysis, it is clear that the overall effect does not have to be monotonic in s^m. Hence we use simulations to illustrate the properties of the probability of interest as a function of the market signal, confirming also the observation in the aforementioned remark[13].



On the y-axis we have the probability of change in rank, and on the x-axis the corresponding market signal. As it can been seen from the graph the total effect is hump-shaped in s^m, it is decreasing as the market signal becomes relatively large and also it is smaller when market conditions are good compared to when they are bad.

In the next section, we find empirical evidence supporting our results. This is done by constructing divisions in which each fund is allocated in accordance with their AUM. Subsequently, we calculate the proportion of funds that changed division from the beginning

[12]This probability is always strictly positive, since we know that $\varphi_0^1(r^1, s^m) \to 1$ as $r^1 \to +\infty$ and $s^m \to -\infty$, or intuitively the return of manager 1 may be much larger than that of manager 2 when they invest in their own projects (either because one has high skill and the other has low or because one is just luckier than the other) and hence this can always lead to a change of ranks.

[13]For this simulation we set the parameters as: $\pi^1 = 0.6$,$\pi^2 = 0.601$, $\alpha^H = 0.16$, $\alpha^L = 0.1$,$\sigma = \nu = 0.35$, $f^1 = f^2 = 0.01$, $\sigma_m \nu_m = 0.25$, $\lambda_1 = \lambda_2 = 0.8$ and $\delta = 0.5$.

of each period to its end. Approximately, this measures the probability to which the above proposition refers.

3.4 Empirics

3.4.1 Data

The data used in this study comes from the Morningstar CISDM database. The time span of our sample is from January 1994 to December 2015. To mitigate survivorship bias we include defunct funds in the sample. We have created a larger group of strategies to accumulate the Morningstar's categories. All fund returns have been converted to USD (U.S dollars) using the exchange rates of each period separately. Observations of performance or assets under management, with more than 30 missing values, have been deleted. All observations are monthly. Our main variable of interest is flows, which gives the proportional in and out flows of the fund with respect to its assets under management. For the market return we consider the S&P 500 and as fund excess returns, the difference of the fund's return with the market. In particular, we use the corresponding Fama-French market factor obtained from the WRDS (or from Kenneth French's website at Darmouth). We also examine the relationship of alpha and beta of a fund as well as their relationship to the flows.

3.4.2 Empirical Evidence

The purpose of this section is to empirically test some of the assumptions as well as the results of our model and show that our model can be empirically supported by data. For simplicity we will use CAPM alpha and beta throughout this section, calculated using a 32 month period (which we will define in this section as *one period*)[14]. Moreover we will refer to the log of the assets of a fund lagged by one period, simply as the fund's *assets*. First of all, our model assumes that investors get a signal about the market (s^m) before everyone else does. This would imply some form of market-timing [15]. We first run the following panel regression, with fixed effects:

$$\text{Beta}_t = \lambda_0 + \lambda_1 r_{m,t} + \lambda_2 \text{Assets}_{t-1} + \lambda_3 \text{Age}_t + d_i + \varepsilon_t$$

where $r_{m,t}$ is the period market return (described above) and d_i corresponds to the fixed effects dummy (although the subscript i for the fund has been suppressed in the rest of the variables). The results are shown below:

The positive and significant coefficient in front of the market return supports our model assumption (as well as with the prediction of Proposition 3 about over-investment), in the

[14]We have also performed robustness check using the 4-factor alphas and betas.

[15]In the empirical literature there have been studies both in favour Chen and Liang (2007) as well as against this finding Franzoni and Schmalz (2017).

Table 3.1 Estimation results : Beta on Market Return.

The baseline model we run is summarised by beta ~ r_m + assets + controls.

Variable	Coefficient	(Std. Err.)
r_m	**0.03256***	(0.01372)
assets	0.01467**	(0.00508)
age	0.00502**	(0.00144)
Intercept	0.02430	(0.08400)

Significance levels : † : 10% $*$: 5% $**$: 1%

sense that it indicates that when markets are bull, it is more likely that managers choose to get higher exposure to the market. This is consistent with what we would observe if managers had market-timing abilities.

Another result of our model is that in equilibrium $l > h$. Given the definition of the cutoff equilibrium strategies described in (5), this leads to: $P(\beta = 1|L) > P(\beta = 1|H)$. If this is the case, we would expect to see in the data that funds with higher alpha, have on average lower betas, i.e they choose to invest on their idiosyncratic project since they benefit both from potential higher returns thanks to their superior alpha as well as from signalling their skill. Indeed this is the case. We are using the following cross-sectional baseline model, for the last date in our data, December 2015[16]:

$$\text{Alpha}_t = \lambda_0 + \lambda_1 \text{Beta}_t + \lambda_2 \text{Assets}_{t-1} + \Lambda_3 \text{Controls} + \varepsilon_t,$$

where controls include the age and the strategy of the fund. As shown in Table 2 the coefficient of interest is negative, suggesting that more skilled managers pick a high beta less often.

Table 3.2 Cross-sectional Regression of Alphas on Betas and controls, $t = 12/2015$.The baseline model we run is summarised by alpha ~ beta + assets + controls.

Variable	Coefficient	(Std. Err.)
beta	**-0.00958****	(0.00085)
assets	0.00006	(0.00016)
age	0.00001	(0.00005)
strategy	0.00003	(0.00009)
Intercept	0.00130	(0.00284)

Significance levels : † : 10% $*$: 5% $**$: 1%

Even more importantly, we want to test the second implication of Proposition 5. That is, we want to test whether the data suggest that the sensitivity of flows to performance is higher

[16]We only include funds that report US dollars as their base currency.

when beta is 0, or consequently is higher when markets are bear than when they are bull. We will measure the fund flows, as in Sirri and Tufano (1998):

$$\text{Flows}_t = \frac{TNA_t - (1 + R_t)TNA_{t-1}}{TNA_{t-1}}$$

where TNA is the total net assets and R is the return of the fund. We will use the simple return of the fund, r_i, as the measure of performance, as in Clifford et al. (2013). We think that this is the most appropriate measure of performance to test the predictions of our model. The following two tables[17] verify the above finding, and support our predictions[18]. First regression is a cross-sectional one for December 2015.

$$\text{AvFlows}_t = \lambda_1 r_{i,t} \cdot \text{Bigbeta}_t + \lambda_2 \text{Assets}_{t-1} + \Lambda_3 \text{Controls} + \varepsilon_t$$

where *AvFlows* is the average flows of the previous period, Bigbeta $= 1_{\{\beta \geq 0.3\}}$, $r_{i,t}$ is the fund's period return and controls include the age, the strategy and the bigbeta dummy of the fund (the intercept λ_0 is just suppressed in the above equation).

Table 3.3 Flows on Performance and Beta, $t = 12/2015$

Variable	Coefficient	(Std. Err.)
r_i·Bigbeta	**-0.12510****	(0.03433)
Bigbeta	0.01204	(0.01522)
assets	-0.01437**	(0.00389)
strategy	0.00352	(0.00231)
age	-0.00133	(0.00120)
Intercept	0.25846**	(0.06906)

Significance levels : † : 10% $*$: 5% $**$: 1%

The second table we are presenting is a panel regression with *fixed effects*, where we regress flows on the interaction of annual fund's performance and market return, including the usual controls. That is, our baseline model is:

$$\text{AvFlows}_t = \lambda_1 r_{i,t} \cdot r_{m,t} + \lambda_2 \text{Assets}_{t-1} + \Lambda_3 \text{Controls} + d_i + \varepsilon_t$$

where controls include the fund's beta and the period return of the market and of the fund itself.

[17] Since the funds selected in our model, are only between $\beta = \{0, 1\}$, thus making the implicit assumption that there is no short-selling of the market, we will exclude all observation with negative β, which are less than 15% of our sample.

[18] This result was only recently documented empirically in a paper by Franzoni and Schmalz (2013).

Table 3.4 Flows on the interaction of Fund Performance and Market Return

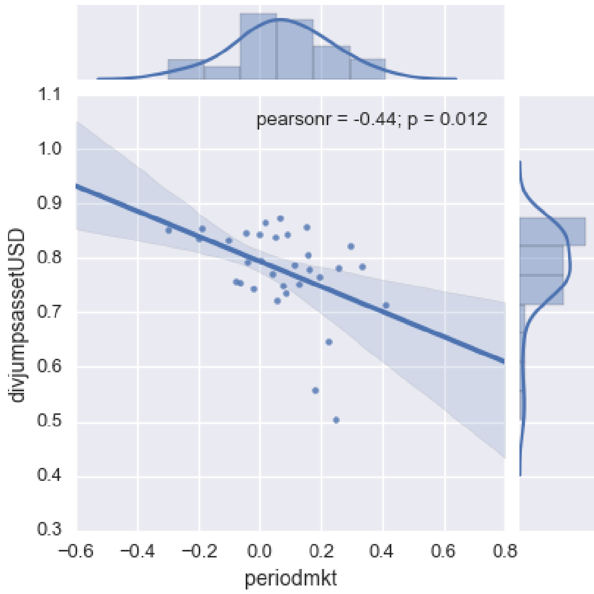
Variable	Coefficient	(Std. Err.)
$r_i \cdot r_m$	**-0.15297****	(0.03697)
beta	0.00423	(0.00648)
r_i	0.07538**	(0.02036)
r_m	0.02828**	(0.00955)
assets	-0.02718**	(0.00224)
Intercept	0.47264**	(0.03964)

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

In both cases we can see that the coefficient of interest is significantly *negative*. The interpretation of these two regressions is the following: the first one shows that funds with higher beta are not judged so much on their performance; that is the higher the beta, the less important the flow performance relationship. On the other hand, the second table supports the statement that the sensitivity of fund flows to performance depends on the state of the market and more specifically it is decreasing on the market return. Under the predictions of our model, these two results are almost equivalent, and we indeed get that the coefficient in both cases is negative and significant, thus supporting one of our main results as well.

Finally, we want to provide some empirical evidence relevant to the discussion on the competition of funds. Namely, we find support for Remark 1, by demonstrating that the probability of changes in the ranking of funds, with respect to their AUM, is higher under adverse market conditions. To achieve this a new variable is constructed. First, the sample is separated in periods of eight months, so that we have thirty periods in total. For each one, seventy divisions (clusters of funds) are created. Funds are allocated in those divisions according to the size of their AUM at the end of each period[19]. Then we define $divjumpassetUSD_t$ as the percentage of funds that changed division from the beginning to the end of period t. We are careful to only compare funds that were active during the whole duration of each period. Also, we only consider the US universe of funds to avoid introducing noise created from fluctuations in the exchange rates.

[19] Our methodology closely follows previous work done by Marathe and Shawky (1999) and Nguyen-Thi-Thanh (2010).



On the y-axis we have our constructed measure of changes between divisions *divjumpassetUSD*$_t$, and on the x-axis the corresponding total return of the market portfolio during the same period. As it can been seen from the graph there appears to be a negative relationship between the two, which is also statistically significant. Note that this is just an indication of the relationship between the rank of funds and the market conditions, under a simple linear regression, and thus it does not capture any second order effects (or a hump-shaped relationship). Hence, our prediction in Remark 1 is only supported by weak evidence, but we believe that there is much more to explore in this direction in the future.

3.5 Extension: Unobservable Investment Decision

In this section, we want to extend our model, and investigate the equilibrium where the investment decision of the fund managers cannot be observed by the investors. In this case, investors use the return of the fund managers' to both update their beliefs about managers skill and to also understand whether or not they invested in their own project. In reality, it is indeed the case that investors do not know exactly the exposure of a fund manager to the systematic risk. Instead they use a history of data of the fund return's comovement with the market return to infer the fund's statistical *beta*. Since the model we are examining here is static, the assumption in this section is that this inference is only made based on the proximity of the market return to the fund's return.

The model considers only one period and it remains the same as before, apart from a few changes outlined below. Firstly, an additional error ϵ has been introduced in order to make the manager's choice of investment unobservable by the investors. (Note, that without this tracking error, investors could perfectly observe the decision of managers based on whether or not $r = m$.) Hence our model becomes:

$$\begin{aligned} r &= (1-\beta)\,a + \beta\,(m+\epsilon) \\ a &\sim \mathcal{N}(\alpha, \sigma^2) \\ m &\sim \mathcal{N}(\mu\,, \sigma_m^2) \\ \epsilon &\sim \mathcal{N}(0\,, \sigma_\epsilon^2) \end{aligned} \tag{3.20}$$

The manager's performance r is a weight average of the return of her idiosyncratic strategy a and that of the market m, and as before we study only the simple binary case where $\beta \in \{0, 1\}$. The rest of the notation and ideas remain unchanged.

The posterior distribution of r, conditional on (α, β, s, s^m) is given by

$$\begin{aligned} r\,|\,\alpha, \beta, s, s^m &\sim \mathcal{N}\Big(\bar{r}(\alpha, \beta, s, s^m),\ \bar{\sigma}^2(\beta)\Big) \\ \bar{r}(\alpha, \beta, s, s^m) &\equiv (1-\beta)[(1-\psi)\alpha + \psi s] \\ &\quad + \beta[(1-\psi_m)\mu + \psi_m s^m] \\ \bar{\sigma}^2(\beta) &\equiv (1-\beta)^2\psi\nu^2 + \beta^2(\psi_m\nu_m^2 + \sigma_\epsilon^2) \end{aligned} \tag{3.21}$$

Our goal is to study whether a monotonic cutoff equilibrium (introduced in the previous sections) exists under this alternative assumption. We believe that only such an equilibrium would be interesting and realistic to serve for further study. We move on to find a closed-form expression for the ex-post reputation φ, which is given by the following lemma.

Lemma 3 *The manager's posterior reputation is given by*

$$\varphi(r, m, s^m) = \left(1 + \frac{1-\pi}{\pi}\frac{\rho(r, L, l(s^m))}{\rho(r, H, h(s^m))}\right)^{-1}, \tag{3.22}$$

where $l(s^m)$ $h(s^m)$ from lemma 1, we have

$$l(s^m) - h(s^m) = \frac{1-\psi}{\psi}(H-L), \tag{3.23}$$

and

$$\rho(r, \alpha, c) = \Phi\left(\frac{r - c(1+\psi) + \alpha\psi}{\nu\sqrt{1+\psi}}\right) \times \frac{\phi\left(\frac{r-\alpha}{\nu\sqrt{\psi(1+\psi)}}\right)}{\nu\sqrt{\psi(1+\psi)}} + \Phi\left(\frac{c-\alpha}{\nu}\right)\frac{\phi\left(\frac{r-m}{\sigma_\epsilon}\right)}{\sigma_\epsilon}, \tag{3.24}$$

Proof: In Appendix B.3. □ Using the above lemma, we can now see whether this model can provide us with an equilibrium where the reputation $\varphi(r, m, s^m)$ is increasing in r. In fact, we get the following proposition:

Proposition 6 *A monotonic equilibrium under unobservable beta does not exist.*

Proof: In Appendix B.3. □

What this proposition shows is that the reputation $\varphi(r, m, s^m)$ cannot always be increasing in r under the assumption that investors do not observe the investment choices. That is to say that the assumption of unobservable investment choice under a static setting can lead us to counterintuitive equilibrium properties. We believe that in future research it could be interesting to study this realistic case under a dynamic setting where the inference of beta will be indeed based on the comovement of the market return with the fund's return.

3.6 Conclusions

The role of financial intermediaries and their characteristics has been greatly explored in the recent empirical literature. In this article, we have developed a theoretical model that describes how the strategic investment decisions of fund managers is influenced by their career concerns. To sum up our argument, these managers will tend to over-invest in market neutral strategies as a way to signal their ability. Moreover, we have described how managers' reputation depends on the market conditions; in particular, we find that the sensitivity of flows to performance is higher in bear markets than in bull markets and we discuss the competition between funds, measured by the changes in their rankings, as a function of the market conditions. Our model entails predictions about some directly observable fund characteristics such as their size and fees, as well as some indirectly observable quantities such as their reputation or their investment behavior depending on their signals. In our empirical section, we have managed to find support for many of the assumptions as well as predictions of our model. Moreover, we have extended our model to include the case when the manager's investment decision is not observable by the investors.

There are many ways forward with this research. The results of this model do not depend on the specific factor which funds use when they are tracking an index; one, may try to apply the same logic in funds that use factors other than the market return and test the corresponding empirical predictions. Also, using a slightly different interpretation of the investor's decision between allocating funds to a manager or to the market, one could think of an investor choosing between an active and a passive fund and use the closed form solution for the fund's size, to see how the relative (total) size of the passive and active funds, depends on the market conditions.

Bibliography

Anthropelos, M. (2017), 'The effect of market power on risk-sharing', *Mathematics and Financial Economics* **11**, 323–368.

Anthropelos, M. and Kardaras, C. (2017), 'Equilibrium in risk-sharing games', *Finance and Stochastics* **21**(3), 815–865.

Anthropelos, M. and Žitković, G. (2010), 'On agent's agreement and partial-equilibrium pricing in incomplete markets', *Mathematical Finance* **20**(3), 411–446.

Babus, A. (2016), 'The formation of financial networks', *RAND Journal of Economics* **47**, 239–272.

Babus, A. and Hu, T.-W. (2016), Endogenous intermediation in over-the-counter markets. Forthcoming in Journal of Financial Economics.

Back, K. (1992), 'Insider trading in continuous time', *The Review of Financial Studies* **5**(3), 387–409.

Back, K., Cao, C. and Willard, G. (2000), 'Imperfect competition among informed traders', *The Journal of Finance* **55**(5), 2117–2155.

Barrieu, P. and El Karoui, N. (2005), 'Inf-convolution of risk measures and optimal risk transfer', *Finance and Stochastics* **9**(2), 269–298.

Berk, J. B. and Green, R. C. (2004), 'Mutual fund flows and performance in rational markets', *Journal of political economy* **112**(6), 1269–1295.

Blume, M. and Keim, D. (2012), Institutional investors and stock market liquidity: Trends and relationships. Working Paper.

Borch, K. (1962), 'Equilibrium in a re-insurance market', *Econometrica* **30**, 424–444.

Brunnermeier, M. and Pedersen, L. (2005), 'Predatory trading', *Journal of Finance* **60**, 1825–1863.

Buhlmann, H. (1984), 'The general economic principle premium', *Astin Bulletin* **14**, 13–21.

Chen, Y. (2015), 'Career concerns and excessive risk taking', *Journal of Economics & Management Strategy* **24**(1), 110–130.

Chen, Y. and Liang, B. (2007), 'Do market timing hedge funds time the market?', *Journal of Financial and Quantitative Analysis* **42**(4), 827–856.

Chevalier, J. and Ellison, G. (1997), 'Risk taking by mutual funds as a response to incentives', *Journal of Political Economy* **105**(6), 1167–1200.

D., D., Scheicher, M. and Vuillemey, G. (2015), 'Central clearing and collateral demand', *Journal of Financial Economics* **116**, 237–256.

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

Duffie, D., Gârleanu, N. and Pedersen, L. (2007), 'Valuation in over-the-counter markets', *The Review of Financial Studies* **20**(6), 1865–1900.

Franzoni, F. and Schmalz, M. C. (2017), 'Fund flows and market states', *The Review of Financial Studies* **30**(8), 2621–2673.

Gibson, S., Singh, R. and Yerramilli, V. (2003), 'The effect of decimalization on the components of the bid-ask spread', *Journal of Financial Intermediation* **12**, 121–148.

Guerrieri, V. and Kondor, P. (2012), 'Fund managers, career concerns, and asset price volatility', *The American Economic Review* **102**(5), 1986–2017.

Hameed, A., Lof, M. and Suominen, M. (2017), Slow trading and stock return predictability. Working paper, available at SSRN: https://ssrn.com/abstract=2671237.

Hendershott, T. and Madhavan, A. (2015), 'Click or call? auction versus search in the over-the-counter market', *Journal of Finance* **70**, 419–447.

Hu, P., Kale, J. R., Pagani, M. and Subramanian, A. (2011), 'Fund flows, performance, managerial career concerns, and risk taking', *Management Science* **57**(4), 628–646.

Huang, J. C., Wei, K. D. and Yan, H. (2012), 'Investor learning and mutual fund flows'.

Ippolito, R. A. (1992), 'Consumer reaction to measures of poor quality: Evidence from the mutual fund industry', *The Journal of Law and Economics* **35**(1), 45–70.

Jouini, E., Schachermayer, W. and Touzi, N. (2008), 'Optimal risk sharing for law invariant monetary utility functions', *Mathematical Finance* **18**(2), 269–292.

Keim, D. and Madhavan, M. (1995), 'The anatomy of the trading process: Empirical evidence on the behavior of institutional traders', *Journal of Financial Economics* **37**, 371–398.

Keim, D. and Madhavan, M. (1996), 'The upstairs markets for large-block transactions. analyses and measurement of price effects', *Review of Financial Studies* **9**, 1–39.

Klemperer, P. and Meyer, M. (1989), 'Supply function equilibria in oligopoly under uncertainty', *Econometrica* **57**(6), 1243–1277.

Kyle, A. (1989), 'Informed speculation with imperfect competition', *Review of Economic Studies* **56**(3), 317–55.

Kyle, A., Obizhaeva, A. and Wang, Y. (2018), 'Smooth trading with overconfidence and market power', *The Review of Economic Studies* **85**(1), 611–662.

Liu, H. and Wang, Y. (2016), 'Market making with asymmetric information and inventory risk', *Journal of Economic Theory* **163**, 73–109.

Ma, L. (2013), 'Mutual fund flows and performance: A survey of empirical findings'.

Madhavan, M. and Cheng, M. (1997), 'In search of liquidity: Block trades in the upstairs and downstairs markets', *Review of Financial Studies* **10**, 175–203.

Magill, M. and Quinzii, M. (2002), *Theory of incomplete markets*, Vol. 1, Mit press.

Malamud, S. and Rostek, M. (2017), 'Decentralized exchange', *Americal Economic Review* **11**, 3320–3362.

Malliaris, S. G. and Yan, H. (2015), 'Reputation concerns and slow-moving capital'.

Marathe, A. and Shawky, H. A. (1999), 'Categorizing mutual funds using clusters', *Advances in Quantitative analysis of Finance and Accounting* **7**(1), 199–204.

Nguyen-Thi-Thanh, H. (2010), 'On the consistency of performance measures for hedge funds', *Journal of Performance Measurement* **14**(2), 1–16.

Robertson, S. (2017), 'Pricing for large positions in contingent claims', *Mathematical Finance* **27**(3), 746–778.

Rostek, M. and Weretka, M. (2008), Thin markets, *in* 'The New Palgrave Dictionary of Economics Online', Eds. S. N. Durlauf and L. E. Blume., Palgrave Macmillan, pp. 307–336.

Rostek, M. and Weretka, M. (2012), 'Price inference in small markets', *Econometrica* **80**(2), 687–711.

Rostek, M. and Weretka, M. (2015), 'Dynamic thin markets', *Review of Financial Studies* **28**, 2946–2992.

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

Vayanos, D. (1999), 'Strategic trading and welfare in a dynamic market', *Review of Economic Studies* **66**(2), 219–54.

Vives, X. (2011), 'Strategic supply function competition with private information', *Econometrica* **79**, 1919–1966.

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

Warther, V. A. (1995), 'Aggregate mutual fund flows and security returns', *Journal of financial economics* **39**(2), 209–235.

Weretka, M. (2011), 'Endogenous market power', *Journal of Economic Theory* **146**(6), 2281 – 2306.

Zawadowski, A. (2013), 'Entangled financial systems', *Review of Financial Studies* **26**(1), 1291–1323.

Appendix A

Appendix on chapter 2

A.1 Appendix: Omitted Proofs

Proof: **[Proof of theorem 2.3.4]**

Let us start with the case where both conditions (2.18) and (2.15) hold. Proposition 2.3.6, together under the validity of (2.16) which is equivalent to (2.18), shows that there exists an extreme linear Nash equilibrium. This is, in fact, unique over *extreme* linear Nash equilibria; indeed, note that (2.18) immediately implies that $\beta_k > 1$. Let us assume that trader $k \in I$ is the *only* trader with beta greater than one, i.e., that $\beta_i \leq 1$ for all $i \in I \setminus \{k\}$. Let $(\theta_i^*)_{i \in I}$ be *any* linear noncompetitive equilibrium in terms of Definition 2.3.1. According to (2.9), $\beta_i \leq 1$ implies that $\theta_i^* \leq \delta_i (1 + \beta_i)_+$, for all $i \in I \setminus \{k\}$. But then,

$$\beta_k \geq 1 + \frac{1}{\delta_k} \sum_{i \in I \setminus \{k\}} \delta_i (1 + \beta_i)_+ \geq 1 + \frac{\theta_{-k}^*}{\delta_k}.$$

By (2.9) again, it follows that $\theta_k^* = \infty$, and applying (2.9) once again, we have $\theta_i^* = \delta_i (1 + \beta_i)_+$, for all $i \in I \setminus \{k\}$, which establishes uniqueness of the extreme Nash equilibrium over *all* possible linear Nash equilibria of Definition 2.3.1.

Having dealt with the case of extreme equilibrium, until the end of the proof we shall assume that (2.15) holds but (2.18) fails. Without loss of generality, let trader $0 \in I$ have the maximal pre-transaction beta: $\beta_i \leq \beta_0$ for all $i \in I \setminus \{0\}$. In view of Lemma 2.3.5, we then have that, necessarily,

$$-1 < \beta_0 < 1 + \frac{1}{\delta_0} \sum_{i \in I \setminus \{0\}} \delta_i (1 + \beta_i)_+. \tag{A.1}$$

Define the set

$$J := \{ i \in I \setminus \{0\} \mid -1 < \beta_i \leq 1 \}.$$

The set $J_0 := J \cup \{0\}$ contains all traders that will eventually submit demand functions with non-zero elasticity. Note that (A.1) implies that $J \neq \emptyset$; indeed, if $J = \emptyset$, then $\beta_0 = 1 - \sum_{i \in I \setminus \{0\}} \beta_i > 1$, and (A.1) would fail, since the quantity at the right-hand side would equal 1.

A Nash equilibrium exists if and only if $\theta_i^* = 0$ holds for all $i \in I \setminus J_0$, while

$$\left(2 + \frac{\theta_I^* - \theta_i^*}{\delta_i}\right) \frac{\theta_i^*}{\theta_I^*} = 1 + \beta_i, \quad \forall i \in J,$$

following from (2.17). Given $\theta_I^* > 0$, θ_i^* for $i \in J$ satisfies the quadratic equation

$$\frac{1}{2}(\theta_i^*)^2 - (\delta_i + \theta_I^*/2)\, \theta_i^* + \delta_i(1 + \beta_i)\theta_I^*/2 = 0, \quad \forall i \in J. \tag{A.2}$$

The discriminant is equal to $\left(\delta_i + \theta_I^*/2\right)^2 - \delta_i(1 + \beta_i)\theta_I^*$, which, since $-1 < \beta_i \leq 1$, is always (regardless of the value of θ_I^*) non-negative. The two roots of equation (A.2) are $\delta_i + \theta_I^*/2 \pm \sqrt{\left(\delta_i + \theta_I^*/2\right)^2 - \delta_i(1 + \beta_i)\theta_I^*}$. Note that since

$$\begin{aligned} \delta_i + \theta_I^*/2 + \sqrt{\left(\delta_i + \theta_I^*/2\right)^2 - \delta_i(1 + \beta_i)\theta_I^*} &\geq \delta_i + \theta_I^*/2 + \sqrt{\left(\delta_i + \theta_I^*/2\right)^2 - 2\delta_i\theta_I^*} \\ &= \delta_i + \theta_I^*/2 + |\theta_I^*/2 - \delta_i| \geq \theta_I^*, \end{aligned}$$

and θ_0^* has to be strictly positive, it holds that $\theta_i^* < \theta_I^*$ for each $i \in J$. Hence, the only root that is acceptable, i.e., the only nonnegative root is

$$\theta_i^* = \delta_i + \theta_I^*/2 - \sqrt{\left(\delta_i + \theta_I^*/2\right)^2 - \delta_i(1 + \beta_i)\theta_I^*}, \quad \forall i \in J.$$

(Recall that our Definition of noncompetitive equilibrium considers linear demand functions with nonpositive slopes.) In other words, and upon defining the function $\phi_i : (0, \infty) \mapsto \mathbb{R}$ via

$$\phi_i(x) := \delta_i + x/2 - \sqrt{(\delta_i + x/2)^2 - \delta_i(1 + \beta_i)x}, \quad x > 0,$$

we should have $\theta_i^* = \phi_i(\theta_I^*)$ for all $i \in J$. The next result gives some necessary properties on ϕ_i for $i \in J$.

□

Lemma A.1.1 *Let $i \in J$. Then, $\phi_i(0+) = 0$, $\phi_i'(0+) = (1 + \beta_i)/2$. Furthermore, ϕ_i is concave, nondecreasing, and such that $\phi_i(\infty) = \delta_i(1 + \beta_i)$.*

Proof: The fact that $\phi_i(0+) = 0$ is immediate. In the special case $\beta_i = 1$, we have $\phi_i(x) = \delta_i + x/2 - |x/2 - \delta_i| = x \wedge (2\delta_i)$ for $x > 0$, and the result is trivial. When $-1 < \beta_i < 1$, ϕ_i is twice continuously differentiable, and an easy calculation gives

$$\phi_i'(x) = \frac{1}{2} - \frac{x/2 - \delta_i\beta_i}{2\sqrt{(\delta_i + x/2)^2 - \delta_i(1+\beta_i)x}}, \quad x > 0,$$

from which it immediately follows that $\phi_i'(0+) = (1+\beta_i)/2$. Furthermore, another easy calculation gives

$$\phi_i''(x) = \frac{-1 + (x/2 - \delta_i\beta_i)^2 \,/\, \big((\delta_i + x/2)^2 - \delta_i(1+\beta_i)x\big)}{2\sqrt{(\delta_i + x/2)^2 - \delta_i(1+\beta_i)x}}, \quad x > 0.$$

Therefore, $\phi_i''(x) < 0$ for all $x > 0$ is equivalent to $(x/2 - \delta_i\beta_i)^2 < (\delta_i + x/2)^2 - \delta_i(1+\beta_i)x$ for all $x > 0$. Calculating the squares and cancelling terms, we obtain $\delta_i^2\beta_i^2 < \delta_i^2$, which is true since $-1 < \beta_i < 1$. Therefore, ϕ_i is concave. Continuing, a straightforward calculation gives

$$\begin{aligned}\frac{x}{2} - \sqrt{(\delta_i + x/2)^2 - \delta_i(1+\beta_i)x} &= \frac{(x/2)^2 - \big((\delta_i + x/2)^2 - \delta_i(1+\beta_i)x\big)}{x/2 + \sqrt{(\delta_i + x/2)^2 - \delta_i(1+\beta_i)x}} \\ &= \frac{-\delta_i^2 + \delta_i\beta_i x}{x/2 + \sqrt{(\delta_i + x/2)^2 - \delta_i(1+\beta_i)x}},\end{aligned}$$

which, as $x \to \infty$, has limit $\delta_i\beta_i$. Therefore, $\phi_i(\infty) = \delta_i(1+\beta_i) > 0$. Since $\phi_i(0) = 0 < \delta_i(1+\beta_i) = \phi_i(\infty)$ and ϕ_i is concave, we conclude that it is nondecrasing. □

Regarding trader $0 \in I$, since $\beta_0 > -1$, in equilibrium we should have

$$\left(2 + \frac{\theta_I^* - \theta_0^*}{\delta_0}\right)\frac{\theta_0^*}{\theta_I^*} = 1 + \beta_0.$$

Note that $\theta_I^* - \theta_0^* = \sum_{i \in J} \theta_i^* = \sum_{i \in J} \phi_i(\theta_I^*)$. Therefore, upon defining

$$\sigma(x) := \sum_{i \in J} \phi_i(x), \quad x > 0,$$

we should have

$$\left(2 + \frac{\sigma(\theta_I^*)}{\delta_0}\right)\frac{\theta_0^*}{\theta_I^*} = 1 + \beta_0,$$

which immediately gives

$$\theta_0^* = \frac{(1+\beta_0)\,\delta_0}{2\delta_0 + \sigma(\theta_I^*)}\theta_I^*,$$

Hence, in equilibrium, the following equation should hold for $\theta_I^* > 0$:

$$\frac{(1+\beta_0)\,\delta_0}{2\delta_0 + \sigma(\theta_I^*)}\theta_I^* + \sigma(\theta_I^*) = \theta_I^*.$$

In other words, at equilibrium θ_I^* should solve the equation

$$\frac{(1+\beta_0)\,\delta_0}{2\delta_0 + \sigma(x)} + \frac{\sigma(x)}{x} = 1, \quad x > 0. \tag{A.3}$$

By Lemma A.1.1, it follows that the left-hand-side of equation (A.3) is decreasing in $x > 0$. Its limit at $x = 0+$ is equal to

$$\frac{1+\beta_0}{2} + \sum_{i\in J}\frac{1+\beta_i}{2} = \frac{|J_0|}{2} + \frac{1}{2}\sum_{i\in J_0}\beta_i.$$

Since $|J_0| \geq 2$ (recall that $J \neq \emptyset$) and $\sum_{i\in J_0}\beta_i \geq 1$ (by definition of J and the fact that $\beta_I = 1$), the above limit is strictly greater than one. It follows that (A.3) will have a (necessarily unique) solution if and only if the limit as $x \to \infty$ of the left-hand-side of (A.3) is strictly less than one. In other words, and since $\sigma(\infty) = \sum_{i\in J}(1+\beta_i)\,\delta_i$, it should hold that

$$(1+\beta_0)\,\delta_0 < 2\delta_0 + \sigma(\infty) = 2\delta_0 + \sum_{i\in J}(1+\beta_i)\,\delta_i,$$

which is exactly (A.1).

The above discussion implies that a unique Nash equilibrium exists under the validity of (2.15) and failure of (2.16), completing the proof of Theorem 2.3.4.

Appendix B

Appendix on chapter 3

B.1 Appendix: Omitted Proofs

Proof: [**Proof of Lemma 1**] Using (3.10) it is easy to argue that both idiosyncratic and index tracking strategies have to be played with positive probability. This is because the effect of the reputation $\varphi_\beta(\cdot)$ on the manager's payoff is bounded, whereas that of current return r is not. But this implies that $\varphi_0(\cdot)$ is calculated using Bayesian updating, and as a result it cannot be a function of r, since in this case r provides no information on the manager's ability α.

Fix s^m, then the manager's expected payoff while investing in an index tracking strategy $\beta = 1$ is not a function of s. On the other hand, her payoff under the idiosyncratic strategy is a function of r. In particular, it follows from the definition of monotonic equilibria that this is increasing in s, which proves that the manager's equilibrium strategy is a cut-off one, as presented in (3.11).

In addition, the indifference condition that defines $h(s^m)$ is

$$\mathbb{E}_r\Big[a + \delta \cdot \lambda \cdot [\varphi_0(r, s^m) \cdot (u^H - u^L) + u^L] \;\Big|\; s = h(s^m),\, \alpha = H\Big]$$
$$= \mathbb{E}_r\Big[m + \delta \cdot \lambda \cdot [\varphi_0(s^m) \cdot (u^H - u^L) + u^L] \;\Big|\; s^m\Big]$$

while the one that defines $l(s^m)$ is

$$\mathbb{E}_r\Big[a + \delta \cdot \lambda \cdot [\varphi_0(r, s^m) \cdot (u^H - u^L) + u^L] \;\Big|\; s = l(s^m),\, \alpha = L\Big]$$
$$= \mathbb{E}_r\Big[m + \delta \cdot \lambda \cdot [\varphi_0(s^m) \cdot (u^H - u^L) + u^L] \;\Big|\; s^m\Big]$$

But the right hand sides of the above two equations are the same, hence the two expressions on the left hand sides are equal. Therefore, the expectations of the two conditional normals that are used in the two left hand sides have to be the same, which implies that

$$(1-\psi)\cdot H \;+\; \psi\cdot h(s^m) \;\;=\;\; (1-\psi)\cdot L \;+\; \psi\cdot l(s^m)$$

from which (3.12) follows. □

Proof: [**Proof of Lemma 2**] The time subscripts are suppressed, when no ambiguity is created. The same is true for the signal s^m in the cutoffs $h(s^m)$ and $l(s^m)$. To find the posterior $\varphi_0(r)$ calculate

$$\mathbb{P}\left(r,\beta=0\,\middle|\,s^m,H\right) \;=\; \mathbb{P}\left(r\,\middle|\beta=0,s^m,H\right)\times\mathbb{P}\left(\beta=0\,\middle|\,s^m,H\right),$$

where

$$\mathbb{P}\left(\beta=0\,\middle|\,s^m,H\right)=\mathbb{P}(s\geq h\,|\,s^m,H)=\Phi\left(-\frac{h-H}{\nu}\right), \tag{B.1}$$

and

$$\mathbb{P}\,(r\,|\,\beta=0,s^m,H)=\int_h^\infty \phi\left(\frac{r-(1-\psi)H-\psi s}{\sqrt{\psi}\,\nu}\right)\times\frac{1}{\sqrt{\psi}\,\nu}\phi\left(\frac{s-H}{\nu}\right)\frac{1/\nu}{\Phi\left(-\frac{h-H}{\nu}\right)}\,\mathrm{d}s$$

Hence, substituting gives that

$$\mathbb{P}\,(r,\beta=0\,|\,s^m,H) \;=\; \int_h^\infty \phi\left(\frac{r-(1-\psi)H-\psi s^i}{\sqrt{\psi}\,\nu}\right)\frac{\phi\left(\frac{s-H}{\nu}\right)}{\sqrt{\psi}\,\nu^2}\,\mathrm{d}s. \tag{B.2}$$

Let $\tilde{s}=(s-H)/\nu$, then the above becomes

$$\begin{aligned}
&\int_{\frac{h-H}{\nu}}^\infty \phi\left(\frac{r-H}{\sqrt{\psi}\,\nu}\;-\;\sqrt{\psi}\,\tilde{s}\right)\frac{\phi(\tilde{s})}{\sqrt{\psi}\,\nu}\,\mathrm{d}\tilde{s}\\
&=\frac{\phi\left(\frac{r-H}{\nu\sqrt{\psi(1+\psi)}}\right)}{\nu\sqrt{\psi(1+\psi)}}\int_{\frac{h-H}{\nu}}^\infty \phi\left(\frac{\tilde{s}-\frac{r-H}{\nu(1+\psi)}}{1/\sqrt{1+\psi}}\right)\sqrt{1+\psi}\,\mathrm{d}\tilde{s}\\
&=\frac{\phi\left(\frac{r-H}{\nu\sqrt{\psi(1+\psi)}}\right)}{\nu\sqrt{\psi(1+\psi)}}\Phi\left(\frac{r-h(1+\psi)+H\psi}{\nu\sqrt{1+\psi}}\right).
\end{aligned} \tag{B.3}$$

Repeat the same process to find $\mathbb{P}\,(r\,|\,\beta=0,s^m,L)$ and observe that it follows from Bayes' rule that

$$\varphi_0(r)=\left(1+\frac{1-\pi}{\pi}\frac{\mathbb{P}\,(r,\beta=0\,|\,s^m,L)}{\mathbb{P}\,(r,\beta=0\,|\,s^m,H)}\right)^{-1}, \tag{B.4}$$

from which the provided formula follows. To derive φ_1 use Bayes' rule to get that

$$\varphi_1 = \left(1 + \frac{1-\pi}{\pi}\frac{\mathbb{P}\,(\beta = 1 \mid s^m, L)}{\mathbb{P}\,(\beta = 1 \mid s^m, H)}\right)^{-1}, \tag{B.5}$$

where $\mathbb{P}\,(\beta = 1 \mid s^m, \alpha) = 1 - \mathbb{P}\,(\beta = 0 \mid s^m, \alpha)$, which has been derived above. □

To prove our existence theorem we need the following three lemmas.

Lemma 4 *If $M(\cdot)$ is the normal hazard function, then for $a \geq b$ we have,*

$$M(a) - M(b) \leq a - b \tag{B.6}$$

Proof: Since the hazard function is a continuous function, we can use the Mean Value Theorem, which says that for any $a > b$ there exists a $\xi \in (a, b)$ such that $M(a) - M(b) = M'(\xi)(a - b)$. Therefore, it is sufficient to prove that $M'(\xi) < 1$ for any ξ. To prove that, note that $M(\cdot)$ is convex, and hence $M'(\cdot)$ is increasing, so it would be sufficient to prove that $\lim_{x\to\infty} M'(x) = 1$. Now we use the following inequality for the normal hazard function. We know that for $x > 0$,

$$x < M(x) < x + \frac{1}{x} \tag{B.7}$$

But this easily implies that $M(x)$ has x as its asymptote as $x \to \infty$ (that is $\lim_{x\to\infty} M(x) - x = 0$). Finally this implies that $\lim_{x\to\infty} M'(x) = 1$ and this completes the proof (note the limit exists because $M'(\cdot)$ is increasing and bounded, as $M'(x) = M(x)(M(x) - x) < 1 + \frac{1}{x^2} < 2$). □

Lemma 5 *The time subscripted is suppressed. A sufficient condition for $\varphi_0(r, s^m)$ to be increasing in the manager's performance r is that*

$$(H - L) \cdot \frac{1-\psi}{\psi} \;\geq\; l(s_1^m) - h(s_1^m). \tag{B.8}$$

Proof: Suppress inputs (r, s^m), and super/sub-scripts. Differentiating gives

$$\frac{\mathrm{d}\varphi}{\mathrm{d}r} = -\frac{\varphi(1-\varphi)}{\nu\sqrt{1+\psi}}\Bigg[-\frac{H-L}{\nu\psi\sqrt{1+\psi}} + M\left(-\frac{r - l(1+\psi) + L\psi}{\nu\sqrt{1+\psi}}\right) - M\left(-\frac{r - h(1+\psi) + H\psi}{\nu\sqrt{1+\psi}}\right)\Bigg] \tag{B.9}$$

Let

$$\begin{aligned} \delta^L &= l(1+\psi) - L\psi \\ \delta^H &= h(1+\psi) - H\psi, \end{aligned} \tag{B.10}$$

then the above is positive if and only if

$$\frac{H-L}{\nu\psi\sqrt{1+\psi}} \geq M\left(\frac{\delta^L - r}{\nu\sqrt{1+\psi}}\right) - M\left(\frac{\delta^H - r}{\nu\sqrt{1+\psi}}\right) \tag{B.11}$$

But using Lemma 4 we see that the right hand side is bounded above by

$$\frac{\delta^L - \delta^H}{\nu\sqrt{1+\psi}} = \frac{(l-h)(1+\psi) + (H-L)\psi}{\nu\sqrt{1+\psi}}. \tag{B.12}$$

Hence, a sufficient condition for the inequality to hold is that

$$\frac{H-L}{\psi} \geq (l-h)(1+\psi) + (H-L)\psi \;\Leftrightarrow\; (H-L)\frac{1-\psi}{\psi} \geq l-h. \tag{B.13}$$

□

Lemma 6 *For $c > 0$, let*

$$\mu(x) = \left(1 + c\frac{\Phi(a_0 + b\,x)}{\Phi(a_1 + b\,x)}\right)^{-1}. \tag{B.14}$$

Suppose $b > 0$, then $\mu'(x) > 0 \Leftrightarrow a_1 < a_0$, whereas $b < 0$ implies that $\mu'(x) > 0 \Leftrightarrow a_1 > a_0$.

Proof: Differentiating gives

$$\mu'(x) = -b\mu(x)[1-\mu(x)] \times [M(-a_0 - b\,x) - M(-a_1 - b\,x)]\,.$$

Then the statement simply follows from the fact that $M(\cdot)$ is increasing . □

Proof: [**Proof of Proposition 1**] Suppress time subscript t. Also suppress the signal s^m in the cutoffs $h(s^m)$ and $l(s^m)$, and in the reputations $\varphi_0(\cdot)$ and $\varphi_1(\cdot)$.

We start by proving existence. As we have argued in Lemma 1, in any monotonic equilibrium the optimal strategy of a high and low type manager is to pick $\beta = 0$ whenever her signal s is above the cutoffs h and l, respectively. In addition, another necessary implication is that h and l satisfy (3.12).

But then Lemma 5 together with (3.12) give that $\varphi_0(r)$ is indeed increasing in r. Hence, the manager's best response to the functional forms of $\varphi_0(\cdot)$ and φ_1 as given in Lemma 2 is to indeed use the cutoff strategies that Lemma 1 describes.

All that remains to prove existence is to show that those cutoffs always exist. To do this note that the manager's payoff maximisation problem when picking the first period's beta is as given in (3.10). Let her expected payoff when picking $\beta = 0$ be denoted by

$$v_0(s,\alpha) \;=\; (1-\psi)\cdot\alpha + \psi\cdot s \;+\; \delta\cdot\lambda\cdot\mathbb{E}_r\left[\log\left(\varphi_0(r)(u^H-u^L)+u^L\right)\,\middle|\, s,\alpha\right],$$

whereas for $\beta = 1$ this becomes

$$v_1 \;=\; (1-\psi_m)\cdot\mu \;+\; \psi_m\cdot s^m \;+\; \delta\cdot\lambda\cdot\log\left(\varphi_1(u^H-u^L)+u^L\right).$$

But then v_1 is bounded, while $v(s,\alpha)$ goes from minus to plus infinity. Hence the manager uses both the low and high beta strategy depending on s. Next, we provide the equation that defines these cutoffs. Rewrite l as a function of h according to

$$l(h) - L = h - H + \frac{H-L}{\psi},$$

and substitute this equality in $\varphi_0(r)$ and φ_1 to obtain the following two functions, in which only h appears out of the two equilibrium cutoffs. Substituting in $\varphi_0(r)$ gives

$$\tilde{\varphi}_0(r,h) \;=\; \left(1+\frac{1-\pi}{\pi}\cdot\rho(r)\cdot\frac{\Phi\left(\frac{r-h\cdot(1+\psi)+H\psi-(H-L)/\psi}{\nu\sqrt{1+\psi}}\right)}{\Phi\left(\frac{r-h\cdot(1+\psi)+H\psi}{\nu\sqrt{1+\psi}}\right)}\right)^{-1}, \tag{B.15}$$

where h is introduced as an input of the function. Similarly, substituting in φ_1 gives

$$\tilde{\varphi}_1(h) \;=\; \left(1+\frac{1-\pi}{\pi}\cdot\frac{\Phi\left(\frac{h-H+(H-L)/\psi}{\nu}\right)}{\Phi\left(\frac{h-H}{\nu}\right)}\right)^{-1} \tag{B.16}$$

Then the cutoff h is given by the high types indifference condition $v_0(h,H) \;=\; v_1$, which using the above notation becomes

$$\begin{aligned}&\delta\cdot\lambda\cdot\int\log\left[\tilde{\varphi}_0(r,h)(u^H-u^L)+u^L\right]\cdot\phi\left(\frac{r-(1-\psi)H-\psi h}{\sqrt{\psi}\nu}\right)\frac{1}{\sqrt{\psi}\nu}\,\mathrm{d}r\\ &\quad=\;\delta\cdot\lambda\cdot\log\left[\tilde{\varphi}_1(h)(u^H-u^L)+u^L\right]+(1-\psi_m)\cdot\mu+\psi_m\cdot s^m-(1-\psi)\cdot H-\psi\cdot h\end{aligned} \tag{B.17}$$

where $\phi(\cdot)$ is the density of the standard normal distribution. To prove existence we demonstrate that (B.17) equation has at least one solution. Let $LHS\,(h)$ denote the left hand side of (B.17), $RHS\,(h)$ its right hand side, and $\Delta(h) = LHS\,(h) - RHS\,(h)$ their difference. Observe that all the parts of the above equation apart from the last line are bounded. As a result,

$$\begin{aligned}\lim_{h\to-\infty}\Delta(h) &= -\infty\\ \lim_{h\to+\infty}\Delta(h) &= +\infty.\end{aligned} \tag{B.18}$$

Then it follows from the continuity of this function that there exists at least one point where $\Delta(h) = 0$. Hence we have proven existence.

Next we show that (3.15) is indeed a sufficient condition for uniqueness. In particular, we will argue that (3.15) implies that $\Delta(h)$ is increasing in h. First, note that $LHS(h)$ is increasing in h, because $\tilde{\varphi}_0(r,h)$ is increasing in both r and h. We have already argued why this is true for r. For h the claim is a direct implication of Lemma 6.

Hence it suffices to identify a condition for $RHS(h)$ to be decreasing. Lemma 6 implies that $\tilde{\varphi}_1(h)$ is increasing in h. This is the opposite monotonicity, however we can use the fact that the following expression has a relatively simple upper bound

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}h}\log\left[\tilde{\varphi}_1(h)(u^H-u^L)+u^L\right] &= \frac{\tilde{\varphi}_1(h)[1-\tilde{\varphi}_1(h)]/\nu}{\tilde{\varphi}_1(h)+\frac{u^L}{u^H-u^L}}\times\left[M\left(-\frac{h-H}{\nu}\right)-M\left(-\frac{l(h)-L}{\nu}\right)\right]\\ &\leq \frac{1}{\nu}\left[M\left(-\frac{h-H}{\nu}\right)-M\left(-\frac{l(h)-L}{\nu}\right)\right]=\frac{1}{\nu}\int_{\frac{L-(1-\psi)H}{\psi}}^{H}M'\left(\frac{x-h}{\nu}\right)\mathrm{d}x\leq\frac{H-L}{\psi\nu^2}\end{aligned} \tag{B.19}$$

Hence, a sufficient condition for the right hand side to be decreasing, which will imply uniqueness, is that

$$\delta\lambda^i\frac{H-L}{\psi\nu^2} \;\leq\; \psi\,,$$

which equivalently gives (3.15). □

Proof: [**Proof of Proposition 2**] We know that $\varphi_0(r, s^m)$ is increasing in r. Hence, it suffices to prove the conjectured result for $r\to-\infty$. The dependence on s^m is suppressed. Let $k = -h(1+\psi)+H\psi$. To find the limit $\lim_{r\to-\infty}\varphi_0(r)$ we first need to calculate.

$$\lim_{r\to-\infty}\frac{\Phi\left(\frac{r+k-(H-L)/\psi}{\nu\sqrt{1+\psi}}\right)}{\Phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)\exp\left(\frac{2(H-L)r-(H^2-L^2)}{2\nu^2\psi(1+\psi)}\right)}. \tag{B.20}$$

Because both the numerator and the denominator go to zero as r goes to minus infinity this limit becomes

$$\frac{e^{\frac{H^2-L^2}{2\nu^2\psi(1+\psi)}}\lim_{r\to-\infty}\frac{\phi\left(\frac{r+k-(H-L)/\psi}{\nu\sqrt{1+\psi}}\right)}{\nu\sqrt{1+\psi}}}{e^{\frac{(H-L)r}{\nu^2\psi(1+\psi)}}\times\left[\Phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)\frac{H-L}{\nu^2\psi(1+\psi)}+\frac{\phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)}{\nu\sqrt{1+\psi}}\right]}.$$

In addition, algebra implies the following simplification

$$\frac{\phi\left(\frac{r+k-(H-L)/\psi}{\nu\sqrt{1+\psi}}\right)}{\phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)}=\exp\left(\frac{2(r+k)-\frac{H-L}{\psi}}{2\nu^2(1+\psi)\psi/(H-L)}\right). \tag{B.21}$$

This in turn gives

$$e^{-\frac{(H-L)r}{\nu^2\psi(1+\psi)}}\frac{\phi\left(\frac{r+k-(H-L)/\psi}{\nu\sqrt{1+\psi}}\right)}{\phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)}=\exp\left(\frac{2k-\frac{H-L}{\psi}}{2\nu^2(1+\psi)\psi/(H-L)}\right).$$

Hence the limit becomes

$$\exp\left(\frac{2k+H+L-\frac{H-L}{\psi}}{2\nu^2(1+\psi)\psi/(H-L)}\right)\times\lim_{r\to-\infty}\left(\frac{\Phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)}{\phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)}\frac{H-L}{\nu\psi\sqrt{1+\psi}}+1\right)^{-1},$$

where

$$\lim_{r\to-\infty}\frac{\Phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)}{\phi\left(\frac{r+k}{\nu\sqrt{1+\psi}}\right)}=\lim_{x\to\infty}\frac{1-\Phi(x)}{\phi(x)}=0 \tag{B.22}$$

Hence, substituting k we obtain that

$$\lim_{r\to-\infty}\varphi_0(r)=\left(1+\frac{1-\pi}{\pi}\exp\left[\left(H-\frac{H-L}{2\psi}-h\right)\frac{H-L}{\psi\nu^2}\right]\right)^{-1}.$$

Next, we want to show that the above is greater than $\varphi_1(r)$ for every h. This holds if and only if

$$\exp\left[\left(H-\frac{H-L}{2\psi}-h\right)\frac{H-L}{\psi\nu^2}\right] < \frac{\Phi\left(\frac{h-H+(H-L)/\psi}{\nu}\right)}{\Phi\left(\frac{h-H}{\nu}\right)} \tag{B.23}$$

which can equivalently be rewritten as

$$\left(H-\frac{H-L}{2\psi}-h\right)\frac{H-L}{\psi\nu^2} < \log\frac{\Phi\left(\frac{h-H+(H-L)/\psi}{\nu}\right)}{\Phi\left(\frac{h-H}{\nu}\right)}. \tag{B.24}$$

Differentiating the left hand side minus the right hand side we get

$$-\frac{H-L}{\psi\nu^2}+\frac{1}{\nu}M\left(\frac{H-h}{\nu}\right)-\frac{1}{\nu}M\left(\frac{H-h}{\nu}-\frac{H-L}{\nu\psi}\right) \;\leq\; -\frac{H-L}{\psi\nu^2}+\frac{H-L}{\psi\nu^2} \;=\; 0 \tag{B.25}$$

Hence it suffices to check that

$$\lim_{h\to-\infty}\frac{\Phi\left(\frac{h-H}{\nu}\right)}{\exp\left(\frac{(H-L)h}{\psi\nu^2}\right)\Phi\left(\frac{h-H+(H-L)/\psi}{\nu}\right)} \leq \exp\left[\left(\frac{H-L}{2\psi}-H\right)\frac{H-L}{\psi\nu^2}\right].$$

Similar argumentation with the above shows that the limit on the left hand side becomes

$$\begin{aligned}
&\lim_{h\to-\infty}\frac{\phi\left(\frac{h-H}{\nu}\right)}{\exp\left(\frac{(H-L)h}{\psi\nu^2}\right)\phi\left(\frac{h-H+(H-L)/\psi}{\nu}\right)}\\
&\qquad= \lim_{h\to-\infty}\exp\left(\frac{2(h-H)+\frac{H-L}{\psi}}{2\nu^2}\,\frac{H-L}{\psi}-\frac{(H-L)h}{\psi\nu^2}\right)\\
&\qquad= \exp\left[\left(\frac{H-L}{2\psi}-H\right)\frac{H-L}{\nu^2\psi}\right].
\end{aligned} \tag{B.26}$$

Hence, the above inequality holds. □

Proof: [**Proof of Proposition 3**] The Input s^m is suppressed. First, note that h is the solution of (B.17), that is the solution of $\Delta(h) = 0$, where $\Delta(h)$ is defined under the equation as the difference of its left hand side from its right hand side. Second, the optimal cutoff under no career concerns for the high type $c(H)$ is the one that corresponds to the solution of this equation for $\delta = 0$, as this corresponds to the case when the next period is irrelevant. Let $h(\delta)$ denote the solution of (B.17) as a function of δ. Then it follows from the implicit function theorem that

$$\frac{\mathrm{d}h(\delta)}{\mathrm{d}\delta} = -\frac{\partial\Delta(h)/\partial\delta}{\partial\Delta(h)/\partial h}\bigg|_{h=h(\delta)}. \tag{B.27}$$

But it follows from the limits calculated in (B.18) that the unique monotonic equilibrium needs to have $\partial\Delta(h)/\partial h > 0$. Moreover, calculating the derivative on the numerator for some generic h gives

$$\frac{\partial\Delta(h)}{\partial\,\delta} = \lambda\,\mathbb{E}_r\Big[\log\big[\tilde{\varphi}_0(r,h)(u^H - u^L) + u^L\big] - \log\big[\tilde{\varphi}_1(h)(u^H - u^L) + u^L\big]\,\Big|\, s = h, H\Big],$$

but it follows from Proposition 2 that this is positive, because the difference inside the expectation is positive for every h. As a result, for every $\delta \geq 0$ we get that $\mathrm{d}h(\delta)/\mathrm{d}\delta < 0$, which through (3.12) implies the same for the cutoff used by the low type.

Finally, note that λ and δ enter (B.17) in exactly the same way, hence the same result can be stated for λ. □

Lemma 7 *In the unique monotonic equilibrium, for every prior reputation $\pi > 1/2$ there exists a lower bound $\bar{s}^m(\pi)$, defined as the solution of $\varphi_1(s^m) = 1/2$, such that for every $s^m > \bar{s}^m$ we have $\varphi_1(s^m) > 1/2$, and $\bar{s}^m(\pi^i)$ is increasing in π.*

In addition, for every $s^m \geq \bar{s}^m(\pi)$ the cutoffs $h(s^m)$ and $l(s^m)$ are increasing in π, and the same is true for the posterior reputations $\varphi_0^(r, s^m)$ and $\varphi_1(s^m)$.

Proof: In the proof of Proposition 1 it has been shown that in the unique monotonic equilibrium there exists $\tilde{\varphi}_1$ such that $\varphi_1(s^m) = \tilde{\varphi}_1[h(s^m)]$, and its functional form is given in (B.16). Moreover, it is an immediate implication of Lemma 6 that this is increasing in h, and it is easy to verify that

$$\lim_{h\to+\infty} \tilde{\varphi}_1(h) = \pi. \tag{B.28}$$

In addition, it follows from (B.17), which defines $h(s^m)$, that

$$(1+\psi)H + \psi h(s^m) + \delta\lambda\log\left(\frac{u^H}{u^L}\right) \geq (1-\psi_m)\mu + \psi_m s^m.$$

This provides a lower bound for $h(s^m)$, which is in an increasing function of s^m, and shows that

$$\lim_{s^m\to+\infty} h(s^m) = +\infty, \tag{B.29}$$

from which the existence of the cutoffs follows. It's monotonicity follows from using the implicit function theorem on the equation that defines it

$$\tilde{\varphi}_1[\pi, h(\bar{s}(\pi))] = 1/2, \tag{B.30}$$

where note that $\tilde{\varphi}_1$ is increasing in both π and h, and it has been argued in Proposition 4 that $h(\cdot)$ is also an increasing function.

For the second statement, it follows from (3.12) that it suffices to prove it for $h(s^m)$. Using the implicit function theorem on (B.17) we get that

$$\frac{\mathrm{d}h}{\mathrm{d}\pi} = -\frac{\partial\Delta/\partial\pi}{\partial\Delta/\partial h}, \tag{B.31}$$

where direct differentiation gives $\partial\Delta/\partial h = \psi > 0$ and that

$$\frac{\partial\Delta}{\partial\pi} = \frac{\delta\,\lambda}{\pi(1-\pi)}\mathbb{E}_r\Big[\frac{\tilde{\varphi}_0(1-\tilde{\varphi}_0)}{\tilde{\varphi}_0 + \frac{u^L}{u^H-u^L}} - \frac{\tilde{\varphi}_1(1-\tilde{\varphi}_1)}{\tilde{\varphi}_1 + \frac{u^L}{u^H-u^L}}\,\Big|\, s = h, H\Big], \tag{B.32}$$

where the inputs r and s^m have been suppressed. Some basic calculus shows that for every $\tilde{\varphi} \in [1/2, 1]$ the ratio

$$\frac{\tilde{\varphi}(1-\tilde{\varphi})}{\tilde{\varphi} + \frac{u^L}{u^H-u^L}} \tag{B.33}$$

is decreasing in $\tilde{\varphi}$. Moreover, we have from Proposition 2 that $\tilde{\varphi}_0(r, h) > \tilde{\varphi}_1(h)$ for every $r \in \mathbb{R}$. But we already showed that $\tilde{\varphi}_1(h) > 1/2$ for every $s^m \geq \bar{s}^m(\pi)$. Hence, we get that $\partial\Delta/\partial\pi < 0$, which implies the second statement.

Finally, the third statement follows trivially from noting that the direct derivative of both the posteriors with respect to π is positive, and the fact that both are increasing in $h(s^m)$, implied by Lemma 6, for which it has already been argued that it is increasing in π. □

Proof: [**Proof of Proposition 5**] First, consider the investment decision of a high type manager, for which the probability of choosing the low beta strategy, conditional on the market signal s^m, is

$$\mathbb{P}(\beta = 0 \,|\, s^m) = \mathbb{P}(s \geq h(s^m) \,|\, s^m) = \mathbb{P}(h^{-1}(s) \geq s^m \,|\, s^m), \tag{B.34}$$

since it was shown in Proposition 4 that $h(\cdot)$ is increasing. Moreover, for given s^m the distribution of m is normal and is given by

$$m \,|\, s^m \sim \mathcal{N}\Big((1-\psi_m)\mu + \psi_m s^m\,,\, \psi_m \nu_m^2\Big). \tag{B.35}$$

Let $\tilde{m} = [m - (1-\psi_m)\mu]/\psi^m$. Then

$$\tilde{m} \,|\, s^m \sim \mathcal{N}\Big(s^m\,,\, \nu_m^2/\psi_m\Big), \tag{B.36}$$

while the ex-ante distribution of s^m is

$$s^m \sim \mathcal{N}(\mu\,,\,\sigma_m^2 + \nu_m^2), \tag{B.37}$$

As a result using again the properties of Bayesian updating with normal distributions we get that

$$s^m \,|\, \tilde{m} \sim \mathcal{N}\left(\tilde{\psi}\mu + (1-\tilde{\psi})\tilde{m}\,,\, \frac{\tilde{\psi}\nu_m^2}{\psi_m}\right), \tag{B.38}$$

where $\tilde{\psi}_m = (\sigma_m^2 + \nu_m^2)/(\sigma_m^2 + \nu_m^2 + \nu_m^2/\psi_m)$. Hence for every $\hat{m}$, m such that $\hat{m} > m$, the distribution of corresponding normal that generates s^m conditional on $\hat{m}$ first order stochastically dominates the one of m. This immediately implies that

$$\mathbb{P}(\beta = 0 \,\big|\, \hat{m}) < \mathbb{P}(\beta = 0 \,\big|\, m). \tag{B.39}$$

Hence under better observed market conditions the manager is less likely to have chosen to invest in her idiosyncratic strategy. The second statement of the proposition follows from noting that

$$\frac{\mathrm{d}\varphi_0(r, s^m)}{\mathrm{d}r} \geq 0 = \frac{\mathrm{d}\varphi_1(s^m)}{\mathrm{d}r}, \tag{B.40}$$

To calculate the left derivative it is more convenient to use the equivalent $\tilde{\varphi}_0$ function from the proof of proposition 1. The derivative of this can be calculated in a manner similar to that used in the proof of Lemma 5 to be

$$\frac{\mathrm{d}\tilde{\varphi}_0(r,h)}{\mathrm{d}r} = \frac{\tilde{\varphi}_0(1-\tilde{\varphi}_0)}{\nu\sqrt{1+\psi}}\left[\frac{H-L}{\nu\psi\sqrt{1+\psi}} - \int_{\underline{x}}^{\overline{x}} M'\left(x + h\sqrt{1+\psi}/\nu\right)\mathrm{d}x\right],$$

where $M(\cdot)$ is the hazard rate of the standard normal distribution,

$$\underline{x} = -\frac{r + H\psi}{\nu\sqrt{1+\psi}} \quad \text{and} \quad \overline{x} = \underline{x} + \frac{(H-L)/\psi}{\nu\sqrt{1+\psi}}. \tag{B.41}$$

Next we want to show that this derivative is decreasing in s^m. This appears in $\tilde{\varphi}_0$ only indirectly through the cutoff $h(s^m)$, which has already been shown to be an increasing function. Hence calculate

$$\frac{\mathrm{d}^2\tilde{\varphi}_0(r,h)}{\mathrm{d}r\mathrm{d}h} = \frac{1-2\tilde{\varphi}_0}{\tilde{\varphi}_0(1-\tilde{\varphi}_0)}\left(\frac{\mathrm{d}\tilde{\varphi}_0(r,h)}{\mathrm{d}r}\right)^2 - \frac{\tilde{\varphi}_0(1-\tilde{\varphi}_0)}{\nu^2}\int_{\underline{x}}^{\overline{x}} M''\left(x + h\sqrt{1+\psi}/\nu\right)\mathrm{d}x\,,$$

the second line of which is always negative, as $M(\cdot)$ is a convex function. The first line is negative as long as $\tilde{\varphi}_0(r,h) > 1/2$. But we have already argued in Proposition 2 that

$\tilde{\varphi}_0(r,h) > \tilde{\varphi}_1(h)$, and in Lemma 7 that there exists lower bound $\bar{s}^m(\pi)$ such that for all $s^m \geq \bar{s}^m(\pi)$ it has to be that $\tilde{\varphi}_1(h) > 1/2$. Moreover, the same Lemma gives that $\bar{s}^m(\pi)$ is an increasing function and it is easy to verify that for bounded m

$$\lim_{\pi \to 1} \mathbb{P}(\phi_1(s^m) < 1/2 \,|\, m) \;=\; 0. \tag{B.42}$$

Hence, indeed $\dfrac{\mathrm{d}\varphi_0(r, s^m)}{\mathrm{d}r}$ is decreasing in s^m, from which the second statement of the proposition also follows.

□

Proof: [**Proof of equation 3.18**] We have:

$$\begin{aligned}\mathbb{P}(\varphi^1 > \varphi^2 \,|\, s^m) \;=\; & \mathbb{P}(\varphi_1^1 > \varphi_1^2 \,|\, s^m)\, \mathbb{P}(1,1 \,|\, s^m) \\ & + \mathbb{P}(\varphi_1^1 > \varphi_0^2 \,|\, s^m)\, \mathbb{P}(1,0 \,|\, s^m) + \mathbb{P}(\varphi_0^1 > \varphi_1^2 \,|\, s^m)\, \mathbb{P}(0,1 \,|\, s^m) \\ & + \mathbb{P}(\varphi_0^1 > \varphi_0^2 \,|\, s^m)\, \mathbb{P}(0,0 \,|\, s^m),\end{aligned} \tag{B.43}$$

It follows immediately from Lemma 7 that $\varphi_1^2 > \varphi_1^1$. Moreover, Proposition 2 gives that $\varphi_0^2 > \varphi_1^2$, hence we also have that $\varphi_0^2 > \varphi_1^1$. As a result the above becomes

$$\mathbb{P}(\varphi^1 > \varphi^2 \,|\, s^m) \;=\; \mathbb{P}(\varphi_0^1 > \varphi_1^2 \,|\, s^m)\, \mathbb{P}(0,1 \,|\, s^m) + \mathbb{P}(\varphi_0^1 > \varphi_0^2 \,|\, s^m)\, \mathbb{P}(0,0 \,|\, s^m), \tag{B.44}$$

we can only be certain about the monotonicity of the probability of both managers investing in their idiosyncratic portfolio which is deceasing given a large s^m. The rest of the terms can not be monotonic as we have observed through simulations.

□

B.2 Appendix: Investment and AUM in the Second Period

Here, we first derive the optimal investment decision of a manager in the second period. Second, we use this to calculate her AUM as a function of her posterior reputation, which we later use in order to derive her continuation payoff from period 2. To avoid repetition we consider the extended model in which there are two fund managers. In this the investor's preferences are given by

$$v(i, z_t^{ij}) = \begin{cases} \exp(z_t^{i1} - \bar{z}) \cdot (1 - f_t^i) \cdot R_t^i & , i = 1, 2 \\ \exp(m_t) & , i = m \end{cases}$$

Hence, in this case there are two independent preference shocks , one for each fund. The results of the baseline can be obtained by setting the fees of the second manager equal to one, which will ensure that no investor will invest in her fund.

We solve the second period backwards by first considering the manager's investment decision when the funds have already been allocated. The manager's expected payoff is

$$\mathbb{E}\left[\log\left(A_2^i f_2^i R_2^i\right) \mid s_2^i, s_2^m, \beta_2^i, \alpha\right] = \log\left(A_2^i f_2^i\right) + \mathbb{E}\left[r_2^i \mid s_2^i, s_2^m, \beta_2^i, \alpha\right]$$

As a result the manager's objective when choosing her investment strategy β_2^i in the second period is to simply maximise the expected return r_2^i. Thus, she invests in her alpha only if

$$\mathbb{E}\left[r_2^i \mid s_2^i, s_2^m, \beta_2^i = 0, \alpha\right] \geq \mathbb{E}\left[r_2^i \mid s_2^i, s_2^m, \beta_2^i = 1, \alpha\right] \tag{B.45}$$

It is known that the posterior distributions of a_2^i and m_2, after conditioning on s_2^i and s_2^m, are also normal distributions with known expected values. Let $\psi = \sigma^2/(\sigma^2 + \nu^2)$ and $\psi_m = \sigma_m^2/(\sigma_m^2 + \nu_m^2)$. Then (B.45) becomes

$$(1 - \psi) \cdot \alpha + \psi \cdot s_2^i \geq (1 - \psi_m) \cdot \mu + \psi_m \cdot s_2^m ,$$

which allows us to derive the manager's optimal investment strategy in the second period. This is a cutoff rule such that she invests in her alpha only if $s_2^i \geq c(\alpha, s_2^m)$, where

$$c(\alpha, s_2^m) = \frac{\psi_m}{\psi} \cdot s_2^m + \frac{1 - \psi_m}{\psi} \cdot \mu - \frac{1 - \psi}{\psi} \cdot \alpha \tag{B.46}$$

Thus, for the same market conditions a high type manager invests relatively more frequently on her alpha in the second period, as $c(H, s_2^m) < c(L, s_2^m)$ implies

$$\mathbb{P}[s_2^i \geq c(H, s_2^m)] > \mathbb{P}[s_2^i \geq c(L, s_2^m)] \;\Rightarrow\; \mathbb{P}(\beta_2^i = 0 \,|\, m_2, \alpha = H) > \mathbb{P}(\beta_2^i = 1 \,|\, m_2, \alpha = L),$$

where the second line is required to infer s_2^m from the realised m_2. We will frequently need to condition expectations with respect to m_t instead of s_t^m, because we do not have some measure of the latter in our data.

An important point that needs to be made is that the cutoffs $c(\alpha, s_2^m)$ are not the optimal ones for the investors. This is because those are risk-neutral, while the managers are risk-averse. Following the same argumentation as above we can show that the optimal cutoff for the investors is

$$c^*(\alpha, s_2^m) = c(\alpha, s_2^m) + \frac{\psi_m \sigma_m^2 - \psi \sigma^2}{2\,\psi}. \tag{B.47}$$

Thus the investor's optimal cutoff is adjusted by a "risk-loving" factor. For example, suppose that $\psi_m \sigma_m^2 > \psi \sigma^2$, that is investing in the market is relatively more risky conditional on the information that the manager has at her disposal when making the decision. Then an investor would require a higher level of confidence on her alpha s_2^i in order to also agree that relying on it is preferable to 'gambling' with r_2^m.

Let u_2^α denote the equilibrium payoff of an investor in the second period, conditional on investing with a manager of type α, but net of his preference shock z_t^{ij} and fees f_2^i. Then this is given by

$$u_2^\alpha = \mathbb{P}[s_2^i \geq c_\alpha(s_2^m)]\,\mathbb{E}[R_2^i \,|\, s_2^i \geq c_\alpha(s_2^m)] + \mathbb{P}[s_2^i \leq c_\alpha(s_2^m)]\,\mathbb{E}[R_2^i \,|\, s_2^i \leq c_\alpha(s_2^m)], \tag{B.48}$$

which has a closed form representation that can be derived using the formulas of the moment generating function of the truncated normal distribution. We avoid providing this here as it does not facilitate the understanding of the model in any meaningful way. However, it is important to point out that when the market's posterior variance $\psi_m \sigma_m^2$ is much bigger than that of the alpha-based strategy $\psi \sigma^2$ then the misalignment between the managers' and the investors' preferences could be so substantial that a low type manager would be preferable simply because she is more reluctant to use her alpha. We exclude that by assuming $u_2^H > u_2^L$, because if the parameters of the model were such that investing in an index tracking strategy was so attractive, then there would be little need for professional investors.

Let φ^i denote the public posterior belief on manager i's ability α^i at the beginning of period two. Then the investor's expected payoff, net of fees and the preferences shock, from opting for fund i is

$$u_2^i = \varphi^i(u_2^H - u_2^L) + u_2^L,$$

and the corresponding actual payoff is $e^{z_t^{ij}}(1 - f_t^i)u_2^i$. In addition, each investor has an outside option, which is to ignore the financial intermediaries and instead invest directly on m_2, which gives expected payoff

$$u^m = \mathbb{E}[\exp(m_t)] = e^{\mu+\sigma_m^2/2}.$$

To avoid repetition note that in a manner similar to the one above we can define

$$u_1^i = \pi^i(u_1^H - u_1^L) + u_1^L,$$

as the expected net payoff of an investor active in the first period. However, in this case the functional form of u_1^α will be completely different, as the cutoffs used by the managers in the first period will be influenced by their career concerns. We will derive those under a market equilibrium in the next subsection.

To ensure that when the lowest preference shocks are realised the investor would rather invest directly in the market we will assume that

$$(1 - f_2^i) \cdot u_2^H < u^m \cdot e^{\bar{z}} \tag{B.49}$$

We are now ready to derive the AUM of fund i in the beginning of period t, as only a function of net expected payoffs and announced fees.

Lemma 8 *In any market equilibrium the AUM of fund i, competing against fund k, in period t is*

$$\left(\frac{(1 - f_t^i)u_t^i}{u^m}\right)^{\lambda^i} \left(1 - \frac{\lambda^i}{\lambda^i + \lambda^k}\left(\frac{(1 - f_t^k)u_t^k}{u^m}\right)^{\lambda^k}\right). \tag{B.50}$$

Proof: To simplify the algebra drop the investor superscript and time subscripts. Also let $\xi^i = \log(1 - f^i)u^i$, $i = 1, 2$ and $\xi^m = \log u^m + \bar{z}$. For an investor to prefer fund 1 to both directly investing in the market and to fund 2, it has to be that

$$\exp(z^1 - \bar{z}) \cdot (1 - f^1) \cdot u^1 \geq u^m \Leftrightarrow z^1 \geq \xi^m - \xi^1$$

and

$$\exp(z^1)(1 - f^1)u^1 \geq \exp(z^2)(1 - f^2)u^2 \Leftrightarrow z^1 + \xi^1 - \xi^2 \geq z^2,$$

respectively. Hence the proportion of the market that fund 1 captures is

$$
\begin{aligned}
&\mathbb{P}\left(z^1 \ \geq \ \xi^m - \xi^1 \ \cap \ z^1 + \xi^1 - \xi^2 \ \geq \ z_2\right) \\
&\quad = \int_{\xi^m - \xi^1}^{\infty} \mathbb{P}\left(z^1 + \xi^1 - \xi^2 \ \geq \ z_2 \,\middle|\, z^1\right) \mathrm{d}\mathbb{P}(z^1) \\
&\quad = \int_{\xi^m - \xi^1}^{\infty} \left(1 - e^{-\lambda^2 (z^1 + \xi^1 - \xi^2)}\right) \lambda^1 e^{-\lambda^1 z^1} \mathrm{d}z^1 \\
&\quad = e^{-\lambda_1 (\xi^m - \xi^1)} - e^{-\lambda^2 (\xi^1 - \xi^2)} \frac{\lambda^1}{\lambda^1 + \lambda^2} e^{-(\lambda^1 + \lambda^2)(\xi^m - \xi^1)} \\
&\quad = \left(\frac{(1 - f^1) u^1}{u^m \cdot e^{\bar{z}}}\right)^{\lambda^1} \cdot \left(1 - \frac{\lambda^1}{\lambda^1 + \lambda^2} \left(\frac{(1 - f^2) u^2}{u^m \cdot e^{\bar{z}}}\right)^{\lambda^2}\right)
\end{aligned}
$$

The proof for fund 2 is equivalent. □

The proof calculates (B.50) as the probability of the intersection of two events. The first is that investor j prefers fund i to fund k. The second is that fund i is preferred to direct investment in the market.

To obtain the assets for the case where there is only one manager set $f^2 = 1$ to get:

$$
\left(\frac{(1 - f_t^i) \cdot u_t^i}{u^m \cdot e^{\bar{z}}}\right)^{\lambda^i} \tag{B.51}
$$

B.3 Appendix: Unobservable Investment Decision

Lemma 9 *For generic a and b:*

$$\phi(a - bx)\phi(x) = \phi\left(x\sqrt{1+b^2} - \frac{a\,b}{\sqrt{1+b^2}}\right)\phi\left(\frac{a}{\sqrt{1+b^2}}\right). \tag{B.52}$$

In addition, for generic $\underline{x}$:

$$\int_{\underline{x}}^{\infty} \phi(a - bx)\phi(x)\,\mathrm{d}x = \Phi\left(\frac{a\,b}{\sqrt{1+b^2}} - \underline{x}\sqrt{1+b^2}\right)\phi\left(\frac{a}{\sqrt{1+b^2}}\right)\frac{1}{\sqrt{1+b^2}}. \tag{B.53}$$

Proof: The first equation follows from

$$\begin{aligned}\phi(a - bx)\phi(x)2\pi \;&=\; \exp\left(-\frac{a^2 - 2abx + b^2x}{2} - \frac{x^2}{2}\right)\\ =\; &\exp\left(-\frac{(1+b^2)x^2 - 2abx + \frac{a^2b^2}{1+b^2}}{2} - \frac{a^2 - \frac{a^2b^2}{1+b^2}}{2}\right)\\ =\; &\exp\left(-\frac{1}{2}\left(\sqrt{1+b^2}\,x - \frac{a\,b}{\sqrt{1+b^2}}\right)^2 - \frac{1}{2}\frac{a^2}{1+b^2}\right).\end{aligned} \tag{B.54}$$

The second equation follows trivially from the first. □

To make the notation more compact write $\bar{r}_H(s)$ and $\bar{r}_L(s)$ instead of $\bar{r}(\alpha, \beta = 0, s, s^m)$ and $\bar{r}_1$ instead of $\bar{r}(\alpha, \beta = 1, s, s^m)$. Similarly, write $\bar{\sigma}^2_\beta$ instead of $\bar{\sigma}^2(\beta)$. Also, let

$$\xi^2 \;\equiv\; \sigma^2_\epsilon \frac{\beta_0^2}{(1-\beta_0)^2}. \tag{B.55}$$

Define the following function

$$\begin{aligned}\rho(r, \alpha, c) \;=\; \Phi&\left(\frac{(\tilde{r} - \alpha)\psi\nu}{\xi\sqrt{\xi^2 + \psi^2\nu^2}} - \sqrt{1 + \frac{\psi^2\nu^2}{\xi^2}}\frac{c - \alpha}{\nu}\right)\\ &\times \frac{\phi\left(\frac{\tilde{r}-\alpha}{\sqrt{\xi^2+\psi^2\nu^2}}\right)}{\sqrt{\xi^2 + \psi^2\nu^2}} + \Phi\left(\frac{c-\alpha}{\nu}\right)\frac{\phi\left(\frac{r-m}{\sigma_\epsilon}\right)}{\sigma_\epsilon},\end{aligned} \tag{B.56}$$

which under the restriction that $\beta_0 = 0$ simplifies to

$$\rho(r,\alpha,c) \quad = \quad \Phi\left(\frac{r - c(1+\psi) + \alpha\psi}{\nu\sqrt{1+\psi}}\right) \times \frac{\phi\left(\frac{r-\alpha}{\nu\sqrt{\psi(1+\psi)}}\right)}{\nu\sqrt{\psi(1+\psi)}} \;+\; \Phi\left(\frac{c-\alpha}{\nu}\right)\frac{\phi\left(\frac{r-m}{\sigma_\epsilon}\right)}{\sigma_\epsilon}, \quad \text{(B.57)}$$

Proof: [**Proof of Lemma 3**] Drop dependence on s^m both in the cutoffs and on the expectations. First, calculate the probability of r and β to be realised under the cutoff h. For $\beta = \beta_0$ define the new random variable

$$\tilde{r} \;\equiv\; \frac{r - \beta_0\, m}{1-\beta_0} \;=\; a + \frac{\beta_0}{1-\beta_0}\,\epsilon, \tag{B.58}$$

for which we have

$$\begin{aligned} \tilde{r}\,|\,s,m \;&\sim\; \mathcal{N}\big((1-\psi)H + \psi s,\; \xi^2\big) \\ \xi^2 \;&\equiv\; \sigma_\epsilon^2\frac{\beta_0^2}{(1-\beta_0)^2} \end{aligned} \tag{B.59}$$

as a result

$$\Pr(\tilde{r},\beta_0\,|\,H) = \int_h^\infty \phi\left(\frac{\tilde{r} - (1-\psi)H - \psi s}{\xi}\right)\frac{1}{\xi}\phi\left(\frac{s-H}{\nu}\right)\frac{1}{\nu}\,\mathrm{d}s \tag{B.60}$$

Below we switch the variable of integration to $\tilde{s} = (s-H)/\nu$ and use the above lemma

$$\begin{aligned} \Pr(\tilde{r},\beta_0\,|\,H) \;&=\; \int_{\frac{h-H}{\nu}}^\infty \phi\left(\frac{\tilde{r}-H}{\xi} - \frac{\psi\nu}{\xi}\tilde{s}\right)\phi(\tilde{s})\frac{1}{\xi}\,\mathrm{d}\tilde{s} \\ &= \Phi\left(\frac{\tilde{r}-H}{\xi}\frac{\psi\nu/\xi}{\sqrt{1+\psi^2\nu^2/\xi^2}} - \sqrt{1+\frac{\psi^2\nu^2}{\xi^2}}\frac{h-H}{\nu}\right) \times \phi\left(\frac{(\tilde{r}-H)/\xi}{\sqrt{1+\psi^2\nu^2/\xi^2}}\right)\frac{1/\xi}{\sqrt{1+\psi^2\nu^2/\xi^2}}, \end{aligned} \tag{B.61}$$

which after some algebra gives that

$$\Pr(\tilde{r},\beta_0\,|\,H) \;=\; \Phi\left(\frac{(\tilde{r}-H)\psi\nu}{\xi\sqrt{\xi^2+\psi^2\nu^2}} - \sqrt{1+\frac{\psi^2\nu^2}{\xi^2}}\frac{h-H}{\nu}\right) \times \phi\left(\frac{\tilde{r}-H}{\sqrt{\xi^2+\psi^2\nu^2}}\right)\frac{1}{\sqrt{\xi^2+\psi^2\nu^2}} \tag{B.62}$$

For $\beta = 1$, we have that $r = m + \epsilon$, hence

$$\Pr(r,\beta_1\,|\,H) \;=\; \phi\left(\frac{r-m}{\sigma_\epsilon}\right)\frac{1}{\sigma_\epsilon}\Phi\left(\frac{h-H}{\nu}\right) \tag{B.63}$$

Hence, we have an expression for

$$\Pr(r\,|\,H) \;=\; \Pr\left(\tilde{r} = \frac{r-\beta_0\, m}{1-\beta_0}, \beta_0 \,\middle|\, H\right) + \Pr\left(r, \beta_1\,|\,H\right) \tag{B.64}$$

$$\Pr\left(\tilde{r} = \frac{r-\beta_0\, m}{1-\beta_0}, \beta_0 \,\middle|\, H\right) = \Pr\left(r, \beta = 0\,|\,H\right) = \frac{\phi\left(\frac{r-H}{\nu\sqrt{\psi(1+\psi)}}\right)}{\nu\sqrt{\psi(1+\psi)}}\Phi\left(\frac{r-h(1+\psi)+H\psi}{\nu\sqrt{1+\psi}}\right). \tag{B.65}$$

he expressions for the low type are identical, therefore it is now trivial to use Bayesian updating to derive the posterior reputation of the manager, and complete the proof of this Lemma. □

Proof: [**Proof of Proposition 6**] We want to investigate if $\phi(r, m, s^m)$ can be always increasing in r. From Lemma 3 it is sufficient to see if ρ can always be decreasing in r, where, $\rho = \frac{\rho_L}{\rho_H}$. From the previous Lemma we get:

$$\rho = \frac{\Phi\left(\frac{r-l(1+\psi)+L\psi}{\nu\sqrt{1+\psi}}\right)\frac{\phi\left(\frac{r-L}{\nu\sqrt{\psi(1+\psi)}}\right)}{\nu\sqrt{\psi(1+\psi)}} + \Phi\left(\frac{l-L}{\nu}\right)\frac{\phi\left(\frac{r-\mu}{\sigma_\varepsilon}\right)}{\sigma_\varepsilon}}{\Phi\left(\frac{r-h(1+\psi)+H\psi}{\nu\sqrt{1+\psi}}\right)\frac{\phi\left(\frac{r-H}{\nu\sqrt{\psi(1+\psi)}}\right)}{\nu\sqrt{\psi(1+\psi)}} + \Phi\left(\frac{h-H}{\nu}\right)\frac{\phi\left(\frac{r-\mu}{\sigma_\varepsilon}\right)}{\sigma_\varepsilon}} \tag{B.66}$$

Firstly, a necessary condition for ρ to be decreasing is: $\nu\sqrt{\psi(1+\psi)} = \sigma_\varepsilon$. After substituting into equation B.66, we get:

$$\rho = \frac{\varepsilon^{A_1 r - C_1}\Phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right) + d_1}{\varepsilon^{A_2 r - C_2}\Phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right) + d_2} \tag{B.67}$$

where $A_1 = \frac{L-m}{\sigma_\varepsilon^2}, C_1 = \frac{L^2-m^2}{2\sigma_\varepsilon^2}, b_1 = l(1+\psi) - L\psi, d_1 = \Phi\left(\frac{l-L}{\nu}\right)$ and similarly for A_2, C_2, b_2, d_2.

Note that $A_1 < A_2$. Then we can take the derivative with respect to r, and get the following proportionality:

$$\rho' \propto e^{A_1 r - C_1} e^{A_2 r - C_2} \Phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right)\Phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right)(A_1 - A_2)$$
$$+ e^{A_1 r - C_1} e^{A_2 r - C_2} \Phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right)\Phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right)\frac{1}{\nu\sqrt{1+\psi}}\left(M\left(\frac{b_1-r}{\nu\sqrt{1+\psi}}\right) - M\left(\frac{b_2-r}{\nu\sqrt{1+\psi}}\right)\right)$$
$$+ d_2\left[\varepsilon^{A_1 r - C_1} A_1 \Phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right) + \varepsilon^{A_1 r - C_1}\frac{1}{\nu\sqrt{1+\psi}}\phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right)\right]$$
$$- d_1[\epsilon^{A_2 r - C_2} A_2 \Phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right) + \epsilon^{A_2 r - C_2}\frac{1}{\nu\sqrt{1+\psi}}\phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right)] \quad \text{(B.68)}$$

Now let P^* denote the first 2 terms of (B.68). Then we would want to check whether the derivative of ρ is negative for every r, m. We have:

$$\frac{\rho'}{\epsilon^{A_1 r - C_1}\epsilon^{A_2 r - C_2}} \propto \frac{P^*}{\epsilon^{A_1 r - C_1}\epsilon^{A_2 r - C_2}} + d_2\left[A_1\frac{\Phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right)}{e^{A_2 r - C_2}} + \frac{1}{\nu\sqrt{1+\psi}}\frac{\phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right)}{e^{A_2 r - C_2}}\right]$$
$$- d_1\left[A_2\frac{\Phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right)}{e^{A_1 - C_1}}\frac{1}{\nu\sqrt{1+\psi}}\frac{\phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right)}{e^{A_1 - C_1}}\right]$$

We take any m such that $A_1, A_2 < 0$. Intuitively, we consider the case of a good realized market. Then $\frac{P^*}{\epsilon^{A_1 r - C_1}\epsilon^{A_2 r - C_2}}$ is finite (as $r \to \infty$) because $\Phi(.) \in [0,1]$ and $M(a) - M(b) \leq a - b$ for $a > b$ (Lemma 4).

We will now show that as $r \to \infty$, the derivative cannot be negative. Indeed, we have that as $\lim_{r\to\infty}\frac{\phi(.)}{e^{A_2 r - C}} = 0$. In addition it is easily shown that, as $r \to +\infty$:

$$\frac{d_2 A_1 \Phi\left(\frac{r-b_1}{\nu\sqrt{1+\psi}}\right)}{e^{A_2 r - C_2}} - \frac{d_1 A_2 \Phi\left(\frac{r-b_2}{\nu\sqrt{1+\psi}}\right)}{e^{A_1 r - C_1}} \sim \frac{d_2 A_1 e^{C_2} e^{r(A_1 - A_2)} - d_1 A_2 e^{C_1}}{e^{rA_1}} \quad \text{(B.69)}$$

where $\sim$ denotes the asymptotic equivalence of the 2 terms.

We know that $A_1 - A_2 < 0$ so $\lim_{r\to\infty} e^{r(A_1-A_2)} = 0$, hence in the limit the above expression is asymptotically equivalent to

$$\frac{0 - d_1 A_2 e^{C_1}}{e^r A_1} \tag{B.70}$$

Finally, we know that $A_1 < 0$ so $e^r A_1 \to 0$ and therefore the whole expression tends to $+\infty$, since is also $A_2 < 0$.

So we can finally conclude that ρ' cannot always be negative, or in other words, a monotonic equilibrium cannot exist.

□