Delegated Blocks

Finance Working Paper N° 860/2022
July 2026

Amil Dasgupta
London School of Economics
Richmond Mathews
University of Maryland

ECGI Working Paper Series in Finance

Delegated Blocks

Working Paper N° 860/2022
July 2026

Amil Dasgupta
Richmond Mathews

Abstract

How does delegation of equity ownership reshape the link between risk sharing and monitoring? We show that delegation contracts optimally separate the two by independently calibrating managerial trading and monitoring incentives. This is achieved despite the inability to commit to, or contract on, these activities directly. When the contract maximizes client surplus, it maps onto activist hedge funds; when it maximizes managers' profits, it maps onto mutual funds that provide risk sharing but no monitoring. Both deliver less monitoring and risk sharing than proprietary blockholding, and block size poorly predicts monitoring.

Amil Dasgupta
London School of Economics

Richmond Mathews
University of Maryland

Delegated Blocks*

Amil Dasgupta† Richmond Mathews‡

This version: July 2026

Abstract

How does delegation of equity ownership reshape the link between risk sharing and monitoring? We show that delegation contracts optimally separate the two by independently calibrating managerial trading and monitoring incentives. This is achieved despite the inability to commit to, or contract on, these activities directly. When the contract maximizes client surplus, it maps onto activist hedge funds; when it maximizes managers' profits, it maps onto mutual funds that provide risk sharing but no monitoring. Both deliver less monitoring and risk sharing than proprietary blockholding, and block size poorly predicts monitoring. Funds that offer monitoring must also offer risk sharing.

*We thank the editor Thomas Philippon, an anonymous AE, and two anonymous referees for constructive feedback. We are grateful to Anat Admati, Patrick Bolton, Francesco Celentano, Pierre Chaigneau, Alvin Chen, Joanne Chen, Peter DeMarzo, Kenneth Deng, Yunzhi Hu, Mark Loewenstein, Nadya Malenko, John Moore, Shri Santosh, Dimitri Vayanos, Vladimir Vladimirov, Josef Zechner and audiences at CUHK, HEC Paris, HKU, HKUST, LSE, Maryland, Melbourne, Nanyang Business School, NUS, SKKU, SMU, UNSW, Vienna Graduate School of Finance, the Cambridge Corporate Finance Theory Symposium (2024), ECGI GCGC (2026), EFA (2023), FIRS (2023), 3rd HKU Summer Finance Conference, IBEO conference (2023), London FIT conference (2024), OxFIT (2024), SHoF Corporate Governance Conference (2024), Vienna Festival of Finance Theory (2024), and WFA (2023) for helpful comments. Dasgupta acknowledges financial support from the ESRC via Research Grant ES/S016686/1.

†London School of Economics and ECGI. Email: a.dasgupta@lse.ac.uk

‡Smith School of Business, University of Maryland at College Park. Email: rmathews@umd.edu

1 Introduction

In modern financial markets, most equity is held by professional asset managers such as mutual funds, hedge funds, and similar intermediaries. These managers serve clients, who are the ultimate beneficial owners, under asset management contracts. This paper asks how the delegation of equity ownership reshapes the link between risk sharing and monitoring, two of the most important consequences of equity holding in financial markets.

When an investor trades on her own account, the link between risk sharing and monitoring is mechanical: an investor who holds more of a firm's equity for risk sharing purposes captures a larger share of the gains from improving the firm, and so monitors more. Delegation breaks this mechanical link because the entity that trades and monitors (the fund) is not the entity that ultimately bears its consequences (the fund's clients), and the relationship between the two is governed by endogenous contracts.

Our central insight is that delegation contracts optimally *separate* risk sharing from monitoring. This separation is optimal no matter who—the fund clients or fund managers—extracts surplus from the asset management relationship, though for different underlying reasons. The common force is the free-riding logic of Grossman and Hart (1980), applied through two distinct lenses.

When the contract is structured to maximize the welfare of fund clients, these clients want the fund to hold more equity than their initial endowment, for risk-sharing purposes, but they value monitoring only at the level justified by that initial endowment. This is because any additional monitoring is capitalized into the price of the shares the fund acquires on their behalf since *other traders* free ride on the fund's monitoring. In this case, we show that the optimal fund trades such that the clients hold their privately preferred risk-sharing position, while a calibrated number of fund managers hold an effective stake equal to the clients' initial endowment and therefore monitor at the clients' privately optimal level. This optimal fund bears a strong resemblance to activist hedge funds, in which managers hold significant stakes in their own funds and monitor intensively.

When the contract is instead structured to maximize the profits of managers, these managers cannot extract fees for monitoring, because monitoring is *non-excludable.* Clients would not pay for a service that they can enjoy if they stay out of the fund. Here, therefore, the challenge is free riding by *the clients themselves.* A profit-maximizing manager therefore finds it unprofitable to bear the cost of managerial skin in the game that would result in monitoring. In this case, we show that the optimal fund still trades such that the clients hold their privately preferred risk sharing position, but the fund managers hold no stake in the fund and undertake no monitoring whatsoever. This optimal fund bears a strong resemblance to mutual funds.

A notable feature of our results is that the optimal contracts achieve separation *despite* the inability to commit to, or contract on, trading and monitoring decisions. Once the fund is formed, fund managers are free to trade and monitor as they like, and no contractual instrument constrains either choice directly. The contract operates *indirectly* through three primitives: the number of fund managers, the fraction of fund assets they hold ("skin in the game"), and an upfront fee. These primitives are chosen so that fund managers, acting purely in their own interest, voluntarily trade to a position at which monitoring and overall ownership coincide with the levels the clients privately prefer (in the case where clients extract surplus) or to the levels that maximize fund managers' profits (in the case where managers extract surplus).

Regardless of whether activist hedge funds or mutual funds arise in equilibrium, risk sharing and monitoring both fall below the levels that would arise in the presence of a proprietary trader with high risk bearing capacity. However, the endogenously optimal asset management structure does bring clear risk-sharing gains and (in some cases) beneficial monitoring relative to the case of autarky.

Model. We develop these results in a CARA-Normal economy. A measure $1-\lambda$ of "financial" investors are experts who trade and monitor on their own behalf and, potentially, on behalf of others; a measure λ of "nonfinancial" investors face high opportunity costs of participating directly in financial markets, but can join a fund managed by financial investors. Nonfinancial investors are endowed with a fraction $\omega < \lambda$ of the firm's equity—strictly less than they would hold under competitive risk sharing—leaving them with material unmet risk-sharing needs.

A fund is represented by a contracting triple (τ, ϕ, f): a measure τ of financial investors become Fund Managers (FMs) and hold a fraction ϕ of the fund's assets, while nonfinancial investors become Fund Clients (FCs), hold the rest of the fund's assets, and pay a per-FC upfront fee f. This is a minimal formulation, since any description of a fund must entail specifying who manages the fund, how the value of the fund is divided between clients and managers, and a means of transferring surplus. The fund cannot commit to a trading strategy or monitoring level, and these decisions are not contractible.

The hedge fund case: fund clients have bargaining power *and* high endowments. We first study the contract that is optimal from the clients' perspective. To do this, we initially compute the privately optimal allocation from the FCs' point of view *as if* they could trade and monitor *and also* commit to specific monitoring and trading strategies. In this optimum, FCs would commit to monitor at a level corresponding to their endowment ω, while committing to hold a larger position in the asset, $\frac{\lambda}{1+\lambda}(1+\omega)$, to attain the maximum benefit from risk sharing while factoring in their collective market power. Then we show that a contract of the form (τ, ϕ, f) can be calibrated to deliver this optimum. This contract ensures that FMs will trade on behalf of the fund to the point where they hold an effective stake equal to the FCs' initial endowment (ω), while the FCs hold their privately preferred risk-sharing position ($\frac{\lambda}{1+\lambda}(1+\omega)$). The fee f compensates the FMs for the monitoring costs they expect to incur and for any shares they reallocate to clients on joining the fund. Monitoring and risk sharing are thereby implemented at the levels of the FC's full commitment optimum, without any party committing to a trading strategy or monitoring level, and without trading or monitoring entering the contract directly.

Evaluated at parameters that correspond to wealthy clients with bargaining power, this optimal contract maps onto the institutional reality of activist hedge funds. The contract calls for a non-trivial mass of managers, substantial managerial co-investment (the FMs' effective stake equals ω), and meaningful monitoring of portfolio firms. Each prediction matches documented features of hedge funds: hedge fund managers self-invest substantially in their own funds (Agarwal, Daniel, and Naik 2009, He and Krishnamurthy 2013), and hedge funds play a far more active role in monitoring portfolio firms than other classes of intermediary (Brav, Jiang, and Kim 2010). Hedge-fund clienteles are themselves drawn from high-net-worth individuals and institutions—the natural empirical analogue of FCs with high ω—who shop among managers and negotiate the terms of their participation, which is exactly the bargaining power assumption underlying this case.

The mutual fund case: fund managers have bargaining power *or* clients have low endowments. We next examine what happens when bargaining power flips, or when client endowments shrink. These two distinct economic forces deliver the same alternative outcome—a mutual fund-like vehicle with a vanishing measure of managers, vanishingly small managerial skin in the game, no monitoring, and risk sharing as the sole client benefit. The first force operates when the contract is structured to maximize the payoff of FMs rather than FCs. In this case, a profit-maximizing manager strips monitoring out of the contract entirely because clients refuse to pay for monitoring, which is a non-excludable benefit they can free-ride on. As a result, offering monitoring is unprofitable. The fund's remaining role is to provide risk sharing, which is an excludable service that the FMs can sell to FCs at their full willingness to pay, and the resulting risk-sharing surplus accrues entirely to the FMs as profits. The second force operates entirely through client characteristics. When client endowments ω shrink toward zero, the privately optimal level of monitoring vanishes, so even the client-optimal contract reduces to a vehicle whose only function is to deliver the clients' preferred risk-sharing position. Both forces deliver the same allocation; what differs is who captures the risk-sharing surplus.

These mutual fund-like outcomes again map well onto institutional reality. Mutual fund managers self-invest very little (Khorana, Servaes, and Wedge 2007), undertake very limited monitoring (Bebchuk, Cohen, and Hirst 2017), and serve clienteles whose primary benefit from joining the fund is diversification rather than governance. Furthermore, the optimal mutual fund in the model employs a vanishing measure of managers while the optimal hedge fund employs a nontrivial measure. Consistent with this, the typical U.S. mutual fund serves orders of magnitude more clients per manager than the typical U.S. hedge fund.

When does each fund type emerge? Combining these results yields a clean two-dimensional taxonomy along the two key factors that the model isolates: client endowments (which proxy for client wealth and direct exposure to the risky asset), and the bargaining power of clients vis-à-vis managers. Hedge fund-like outcomes with positive monitoring, substantial managerial co-investment, and a non-trivial mass of managers arise only in the corner of the parameter space where *both* factors take high values, i.e., where high-endowment clients have high bargaining power. In every other configuration the model predicts a mutual fund-like outcome, for different economic reasons depending on which factor is low. The asymmetry is informative. Hedge fund-like governance is a *fragile* outcome that requires the simultaneous presence of two specific conditions and breaks down whenever either is absent. The mutual fund-like outcome, by contrast, is *overdetermined*, arising whenever *either* condition fails. This asymmetry helps rationalize the empirical fact that mutual funds vastly outweigh hedge funds in number and in aggregate assets under management, and that activism remains the preserve of a comparatively small slice of the asset management industry.

Complementarity of investor types. We next ask whether financial experts could form a fund that monitors without involving nonfinancial clients. The answer, we show, is no. If any subset of fund members is expected to pay monitoring costs, they will individually prefer not to join the fund since they can free ride on its monitoring efforts. Thus, the only stable arrangement is one in which there is no monitoring. The endogenous formation of monitoring vehicles therefore requires two *complementary* groups of investors: one group that can be charged for an excludable service (risk sharing), making it worthwhile for the other group to bear the cost of a non-excludable public good (monitoring). While this complementarity is necessary, it is not sufficient for monitoring to occur, since mutual-fund-like vehicles bring both investor types together but supply no monitoring at all.

Relationship to the proprietary blocks benchmark. The closest direct benchmark for our analysis is the seminal work of Admati, Pfleiderer, and Zechner (1994, APZ hereafter) on proprietary blocks. APZ analyze an exogenously large investor with high risk-bearing capacity in a setting without contracting, and show that the proprietary blockholder trades to the competitive risk-sharing allocation and monitors at the corresponding level. The result is driven by an endowment effect: from any holding strictly below the competitive level, the proprietary blockholder is tempted to buy a bit more, and this pushes her all the way to the competitive allocation, eroding her ability to exploit her market power. Under optimal delegation, by contrast, the contract effectively enables commitment to a trading strategy that exploits the clients' collective market power, so the fund holds strictly less than the APZ proprietary block of identical risk-bearing capacity (to avoid excessive price impact). Furthermore, monitoring is determined by the FMs' calibrated stake rather than by overall fund holdings, and is therefore strictly below the APZ level. Both differences are direct consequences of separation, and they hold whether the fund is a hedge fund or a mutual fund. Section 4 isolates the two underlying frictions—the separation of ownership from control, and client free-riding—that drive these differences, and presents a full welfare comparison across the two frameworks.

Other realistic features of asset management firms. Finally, in Section 6 we discuss how our parsimonious framework speaks to salient real-world features of the asset management industry that the model omits. First, because the hedge fund is the only vehicle that supplies monitoring—a costly, non-excludable benefit—contractual features such as lockups and convex carry can be understood as instruments that preserve monitoring in richer environments. We argue that lockups can be viewed as a way to prevent clients from redeeming before monitoring costs are incurred, and show that convex carry can substitute for scarce managerial risk-bearing capacity. Second, we note that flow-performance relationships may sometimes affect equilibrium monitoring incentives, particularly for mutual funds. Finally, although our taxonomy delivers mutually exclusive conditions for hedge fund and mutual fund-like outcomes, the real-life coexistence of the two is consistent with our framework once market segmentation is recognized: if hedge funds pursue assets and strategies unavailable to mutual funds, our model applies within each segment, and free-riding continues to limit fund formation, size, and monitoring.

Relationship to the existing literature. Our paper relates most directly to APZ and papers that extend their findings. We present a detailed comparison of our results to the APZ proprietary blocks benchmark in Section 4. DeMarzo and Urosevic (2006) extend APZ to a fully dynamic setting, while Marinovic and Varas (2021) incorporate private information; both retain the focus on proprietary blocks.

More broadly, our work connects to several literatures. At the most basic level, our paper contributes to the theoretical literature on blockholder monitoring, surveyed by Edmans and Holderness (2017). Several papers within this literature take block size as given (starting with Shleifer and Vishny 1986), while others model the endogenous emergence of blocks through market microstructure frictions (starting with Kyle and Vila 1991). Our analysis differs from all of these by explicitly modeling the emergence of *delegated* equity blocks through optimal contracting. Bolton and von Thadden (1998) also assume fully transparent financial markets but focus purely on proprietary blocks.

A growing theoretical literature takes the delegated nature of equity ownership seriously.[1] Several papers (e.g., Dasgupta and Piacentino 2015, Piacentino 2019, Dasgupta and Maug 2023) highlight the negative implications of agency frictions on monitoring at portfolio firms. Brav, Dasgupta, and Mathews (2022) show a positive effect of delegation in that the resulting flow incentives can enable funds with relatively small positions to act collectively to monitor firms. However, none of these papers endogenize the emergence of delegated blockholders or the size of their blocks.

Finally, our paper is related to the literature on the endogenous emergence of financial intermediaries (starting with Diamond and Dybvig 1983), as well as optimal contracting in delegated portfolio management (starting with Bhattacharya and Pfleiderer 1985). Relative

[1] See, e.g., Dasgupta, Fos, and Sautner (2021) for an overview of the theory and evidence. Bolton and Rosenthal (2019) document a transition from proprietary blocks to delegated blocks over time for an important class of US firms.

to the former, which has focused on banking, we consider the emergence of asset managers. Relative to the latter, which considers optimal contracting with respect to trading, we incorporate monitoring considerations as well.

2 Delegated blocks: Model and main results

2.1 Environment

Consider a financial market with a single firm that has one infinitely divisible equity share outstanding, and a risk-free asset in perfectly elastic supply whose gross return is normalized to unity. There is a unit continuum of investors who have CARA utility, each with risk tolerance ρ.[2]

A measure $1-\lambda$ of these investors are "financial" investors, i.e., experts who specialize in trading and monitoring. The remaining $\lambda < \frac{1}{2}$ measure are "nonfinancial" investors who are otherwise occupied and face a high opportunity cost of participating directly in financial markets to trade or monitor.[3] Financial investors are endowed with $1-\omega$ shares, distributed equally among themselves. Nonfinancial investors are endowed with $\omega \in (0, \lambda)$ shares, distributed equally among themselves.

There are three dates. At date 1, potentially many rounds of trading can occur in a Walrasian market. In any given round, investors submit demand functions and a market-clearing price is determined. At date 2, monitoring can occur: at a cost of $c(m)$, where $c'(\cdot) > 0$, $c''(\cdot) > 0$, and $c(0) = 0$, a monitoring entity can exert effort $m \geq 0$ to generate a final equity payoff distributed according to $N(\mu(m), \sigma^2)$, where $\mu'(\cdot) > 0$ and $\mu''(\cdot) \leq 0$. At date 3, all payoffs are publicly realized.

Aggregate risk tolerance. In a Walrasian CARA-Normal market with symmetric information, at any price P, each competitive investor has demand of $\rho\frac{\mu(m)-P}{\sigma^2}$. Thus the total demand of a measure x of small competitive investors is $\rho x\frac{\mu(m)-P}{\sigma^2}$, which is equivalent to the

[2]Risk tolerance is the inverse of the coefficient of absolute risk aversion.

[3]In Section 5.2, we derive a lower bound on the opportunity cost that supports our results.

demand of a single competitive investor with risk tolerance ρx. In other words, the aggregate risk tolerance of a given measure of small competitive investors is proportional to their measure. Since nonfinancial investors' aggregate risk tolerance, $\rho\lambda$, represents a fraction λ of the total risk bearing capacity in the economy while their endowment is $\omega < \lambda$, they hold less of the risky asset than their desired level under competitive risk sharing.

2.2 Fund formation

Nonfinancial investors can access financial markets by joining a fund. A fund is formed at the beginning of the game if a measure of nonfinancial investors and a measure of financial investors sign an asset management contract. The nonfinancial investors who join become Fund Clients (FCs), while the financial investors who join become Fund Managers (FMs).[4] Investors agree to join only if it is incentive compatible for them to do so. The contract requires that the fund become the sole trading vehicle for both FMs and FCs, who coinvest by contributing their initial endowments as well as possible cash injections. The FMs then make trading and monitoring decisions on behalf of the fund based on their contractual incentives.

We assume throughout that there is a small cost, $\epsilon > 0$, required to form a fund. This assumption matches the operational reality of fund formation: launching any investment vehicle entails expenditures on legal and regulatory documentation, registration and compliance, and on custody and administrative infrastructure, none of which can be made arbitrarily small. We fully discuss the role played by ϵ in Section 5.1.

Crucially, we assume that FMs cannot commit to a given trading strategy or monitoring level up front, and that monitoring and trading are not contractible. These assumptions mirror those in the existing literature on proprietary blocks (APZ, DeMarzo and Urosevic 2006) and ensure that our results do not rely on artificial commitment power.

A fund contract specifies three parameters: a measure $\tau \in (0, 1 - \lambda)$ of FMs, a skin in the game parameter $\phi \in [0, 1]$ specifying the FMs' share of the fund's assets, and an up-front fee f paid by each FC to join the fund and shared among the FMs. A fund can thus be

[4] We show in Section 3 that no funds would form without the participation of nonfinancial investors.

represented as a contracting triple (τ, ϕ, f). This formulation is a minimal contracting space given the structure of the model, since any specification of a fund must entail deciding who manages the fund and how ownership of fund assets is split, and must provide a means of transferring surplus.[5]

Our setup captures key aspects of real-world financial markets. In such markets there are professional asset managers (FMs in our model), indirect investors who invest only via those managers (FCs), and investors who trade directly on their own account (financial investors who do not become FMs).

Trading and monitoring within the fund. Since FMs undertake trading and monitoring on behalf of the fund, decisions are taken by a measure τ of financial investors who have a final ownership stake of $\phi\alpha^D$, where α^D is the fund's total holding of the risky asset.

With respect to trading, if each FM trades according to their individual risk tolerance, their aggregate trade can be represented as that of a single competitive investor with risk tolerance $\rho\tau$. With respect to monitoring, if FMs behaved as isolated members of a continuum, complete free-riding would arise and no monitoring would occur. However, fund managers in the real world operate inside fund management companies, where the scope for within-firm coordination likely allows for a level of cooperation not possible among dispersed investors. To reflect what may be feasible within an asset management firm, we assume that FMs make monitoring decisions as a group based on their aggregate contractual incentives once they join the fund, i.e., FMs choose the level of monitoring as would a single investor with a final block of size $\phi\alpha^D$. Despite biasing our analysis in *favor* of monitoring, we show below that we *never* achieve as much monitoring as under proprietary blockholding, and monitoring can sometimes be zero even with this assumption (see section 2.4). It is also important to note that the decision to join the fund remains individually incentive compatible for each financial investor.

[5]Later, in section 6.1.2, we also consider contracts where it is not possible to choose τ freely.

Roadmap A key factor that determines what types of funds emerge from the interaction between fund clients and fund managers relates to who can extract surplus from the relationship. There are many factors that can affect this. For example, if fund clients have access to gate keepers (e.g., investment consultants, wealth management offices of private banks) who can negotiate on their behalf with fund managers, or if the fund management industry is fragmented, then fund clients may be able to extract surplus. Alternatively, if the fund management industry is made up of a small number of large players, or has gate keepers of its own (e.g., defined contribution plan sponsors, fund platforms or supermarkets) then fund managers are likely to be able to extract surplus. Differences in sophistication, skill, and wealth can also affect the distribution of surplus.

For expositional purposes we refer to cases where fund clients (managers) can extract surplus as those where fund clients (managers) "have bargaining power." First, in Section 2.3, we derive the optimal fund when fund clients have bargaining power and show how it maps onto the salient institutional features of activist hedge funds. Second, in Section 2.4, we observe that real-world fund structures also include vehicles that look very different from this hedge-fund characterization—mutual funds in particular—and show that our framework can rationalize "mutual fund-like" outcomes through two distinct economic forces: managerial bargaining power and vanishingly small client endowments. Both forces deliver the same allocation, making mutual funds a robust prediction of the framework. We accompany each formal characterization with a discussion of the real-world salience of the resulting fund. We then provide a unified two-dimensional taxonomy of when each fund type emerges.

2.3 Optimal funds when fund clients have bargaining power

We first characterize optimal fund formation under the assumption that fund clients (FCs)—the nonfinancial investors—have full bargaining power in designing the fund contract. Under this assumption, the optimal contract maximizes the FCs' aggregate payoff subject to the FMs' individual participation constraints. As we shall see, this case is most naturally interpreted as the formation of a hedge fund.

For expositional clarity, we first carry out the analysis *as if* the fund formation cost ϵ is zero, deriving the optimal contract and the FCs' associated payoff under this simplification. We then show later in this section how outcomes would adjust with $\epsilon > 0$. We also discuss the role of ϵ more generally in Section 5.1.

2.3.1 The clients' full-commitment benchmark

Here we characterize what the FCs would *like* to achieve even when not constrained by any particular contractual form. Suppose, counterfactually, that the FCs could act as a single strategic investor and could commit ex ante to both a given monitoring level and a single round of trade. If they trade to a final stake of α, they face a price of $\mu(m) - \frac{1-\alpha}{\rho(1-\lambda)}\sigma^2$. Their joint optimization problem is

$$\max_{m,\alpha}\ \alpha\mu(m) - c(m) - \frac{1}{2\rho\lambda}\alpha^2\sigma^2 - (\alpha-\omega)\left(\mu(m) - \frac{1-\alpha}{\rho(1-\lambda)}\sigma^2\right).$$

Lemma 1. *The FCs' full-commitment optimum has an optimal monitoring level m^C implicitly defined by $\omega = \frac{c'(m^C)}{\mu'(m^C)}$, and an optimal final stake of $\alpha^C \equiv \frac{\lambda(1+\omega)}{1+\lambda}$.*

This result reveals the core tension that the optimal contract must resolve. Since $\omega < \lambda$, the optimal stake satisfies $\omega < \alpha^C < \lambda$: the FCs want to increase their stake above their endowment to enhance risk sharing (so $\alpha^C > \omega$), but avoid trading all the way to their competitive allocation to exploit their collective market power (so $\alpha^C < \lambda$), i.e., they optimize with respect to price impact.

Strikingly, the optimal monitoring level does *not* depend on the FCs' final stake α. By analogy to the standard first-order condition for monitoring (which equates the stake to the ratio of marginal monitoring cost to marginal monitoring benefit), it is clear that FCs desire monitoring "as if" their ownership equals their initial endowment ω, not their final holding α^C. The intuition is related to the free-rider logic of Grossman and Hart (1980): any monitoring beyond the level justified by the FCs' initial endowment gets fully capitalized into the price of shares they acquire due to free riding by other investors, so the incremental monitoring generates no net benefit for FCs. Thus:

Remark 1. FCs want to hold more than their initial endowment for risk-sharing purposes but want monitoring to occur only at the level justified by their initial endowment. Risk sharing and monitoring motives call for different levels of ownership.

Under proprietary blockholding, these two objectives cannot be independently satisfied, because the blockholder's stake simultaneously determines both risk sharing and monitoring. We discuss the proprietary blockholding case, as analyzed by APZ, in Section 4. The key question for a world with delegated blockholding is whether a delegation contract can achieve both objectives simultaneously, despite the inability to commit to, or contract on, trading and monitoring. Let

$$\Pi_{FC}^{C} \equiv \alpha^{C}\mu(m^{C}) - c(m^{C}) - \frac{1}{2\rho\lambda}(\alpha^{C})^{2}\sigma^{2} - (\alpha^{C} - \omega)\left(\mu(m^{C}) - \frac{1-\alpha^{C}}{\rho(1-\lambda)}\sigma^{2}\right)$$

denote the FCs' aggregate equilibrium payoff at their full-commitment optimum.

2.3.2 Fund feasibility

Even if a fund contract can be designed to deliver the optimal payoff to FCs, it may not be feasible because individual nonfinancial investors may prefer to stay outside the fund and free-ride on its monitoring.

Lemma 2. *There exists $\hat{\omega} \in (0, \lambda)$ such that for $\omega \leq \hat{\omega}$, nonfinancial investors will join a fund that delivers an aggregate FC payoff of Π_{FC}^{C}, while for $\omega > \hat{\omega}$ they will not.*

Funds provide two benefits to potential FCs: an *excludable* benefit via risk sharing (which cannot be enjoyed without joining the fund) and a *nonexcludable* benefit via monitoring (which can be enjoyed without joining). If the endowment ω is sufficiently close to the competitive risk-sharing level, risk-sharing benefits are small, and individual nonfinancial investors prefer to defect and enjoy monitoring benefits for free. A fund that delivers a payoff of Π_{FC}^{C} to FCs can arise only when nonfinancial investors have sufficient unmet risk-sharing needs—that is, when their initial exposure to the risky asset falls far enough below their desired level under competitive risk sharing that the excludable benefits of joining a fund outweigh the temptation to free-ride on its monitoring. The condition is best understood not as a restriction to investors with small endowments per se, but as a restriction to investors whose existing portfolios leave material risk-sharing gains on the table.

2.3.3 Equilibrium trading by the fund

We now characterize the fund's equilibrium trading behavior for a given contract (τ, ϕ, f). If α^D is the fund's total ownership upon entering date 2, the FMs hold an effective stake of $\phi\alpha^D$. Their optimal monitoring intensity solves

$$m^D(\alpha^D) = \arg\max_m \Psi^D(\alpha^D),$$

$$\text{where } \Psi^D(\alpha^D) = \phi\alpha^D\mu(m^D) - c(m^D) - \frac{1}{2\rho\tau}\phi^2(\alpha^D)^2\sigma^2 \tag{1}$$

is the certainty equivalent for the FMs holding an effective stake of $\phi\alpha^D$. The solution is given implicitly by $\phi\alpha^D = \frac{c'(m^D(\alpha^D))}{\mu'(m^D(\alpha^D))}$.

The formation of the fund reduces the mass of competitive price-taking investors from $1-\lambda$ to $1-\lambda-\tau$. If competitive investors expect the fund to hold α^D, the market clearing price is

$$P^D(\alpha^D) = \mu(m^D(\alpha^D)) - \frac{1-\alpha^D}{\rho(1-\lambda-\tau)}\sigma^2. \tag{2}$$

Since FMs cannot commit to a final round of trade, we focus on *globally stable allocations*, initially introduced by APZ, adapted for our delegated fund setting.[6]

Definition 1. An allocation α_G^D is globally stable iff (i)

$$\alpha_G^D \in \arg\max_\alpha \Psi^D(\alpha) - \Psi^D(\alpha_G^D) - \phi(\alpha - \alpha_G^D)P^D(\alpha_G^D),$$

and (ii) for every $\omega \in [0,1]$ such that $\omega \neq \alpha_G^D$,

$$\Psi^D(\alpha_G^D) - \Psi^D(\omega) - \phi(\alpha_G^D - \omega)P^D(\alpha_G^D) > 0.$$

[6]While the concept of global stability is static, its relevance has been confirmed in a fully dynamic setting by DeMarzo and Urosevic (2006).

In words, this means that: (i) once a globally stable allocation is reached, the FMs will not wish to trade away from it at current prices; and (ii) FMs are willing to trade to the globally stable allocation from any other position at prices consistent with the globally stable allocation.

Lemma 3. *As long as $\Psi^D(\cdot)$ is strictly concave, there exists a unique globally stable allocation*

$$\alpha_G^D = \frac{\tau/\phi}{\tau/\phi + 1 - \lambda - \tau}. \tag{3}$$

The fund's equilibrium stake depends on the FM and FC population sizes (τ and λ) and the skin in the game parameter (ϕ). The expression corresponds to the competitive equilibrium allocation in a market with two traders: one with risk tolerance $\rho(1 - \lambda - \tau)$ (the non-FM financial investors) and one with risk tolerance $\rho\tau/\phi$ (the fund). The fund's effective risk tolerance is $\rho\tau/\phi$ because the τ FMs (with aggregate risk tolerance $\rho\tau$) are exposed to only a fraction ϕ of the fund's holdings.

Notably, this allocation does not generally deliver perfect risk sharing among all investors, which would require $\alpha_G^D = \tau + \lambda$. The deviation arises because only a measure $\tau < \tau + \lambda$ of investors make decisions on behalf of the whole fund, and those investors are exposed to only a fraction ϕ of the fund's holdings.

We assume for the remainder of the analysis that $\Psi^D(\cdot)$ is strictly concave, so that a globally stable allocation exists.[7] We now characterize how risk is shared among financial investors.

Lemma 4. *The FMs in the fund and the outside financial investors end up with identical effective per-investor holdings of the risky asset.*

This result shows that there is perfect risk sharing among all $1-\lambda$ financial investors, whether inside or outside the fund, over the part of the risky asset not effectively held by FCs. Multiple trading opportunities combined with the inability to commit erode the strategic advantage of FMs, who trade to perfect risk sharing with outside financial investors.

[7]Between the strict concavity of μ and the strict convexity of c, the conditions for strict concavity of Ψ^D are largely ensured. Details can be found in the Internet Appendix, Section 2.

2.3.4 The optimal delegation contract: The separation principle

We now derive the optimal fund contract.

Proposition 1. *For $\omega \leq \hat{\omega}$, it is feasible to form an optimal fund that achieves an aggregate payoff of Π^C_{FC} for the FCs. It is characterized by:*

1. *a mass of FMs $\tau^* = \frac{(1-\lambda^2)\omega}{1-\lambda\omega}$,*
2. *a skin in the game parameter $\phi^* = \frac{(1+\lambda)\omega}{2\lambda\omega+\lambda+\omega}$, and*
3. *a fee*
$$f^* = \frac{1}{\lambda}\left[c(m^C) + P^{D*}(\alpha^{D*}_G)\left((1-\phi^*)\left(\omega + \tau^*\frac{1-\omega}{1-\lambda}\right) - \omega\right)\right] \tag{4}$$
where the superscript $D*$ indicates that the associated function or variable is evaluated at ϕ^* and τ^*.

The proof proceeds in several steps. First, we solve for τ^* and ϕ^* such that: (1) the FMs' equilibrium monitoring effort equals the FCs' optimal level m^C, which requires the FMs' effective stake $\phi^*\alpha^{D*}_G$ to equal ω; and (2) the FCs' effective stake $(1-\phi^*)\alpha^{D*}_G$ equals their optimal stake α^C. Next, we set f^* to just satisfy the participation constraint of individual FMs. Finally, we verify that this fee makes the FCs' aggregate payoff coincide with Π^C_{FC}.

The separation principle. The mechanism through which the contract separates monitoring from ownership is summarized in the following remark.

Remark 2. The delegation of blockholding breaks the link between risk sharing by FCs and the monitoring that occurs on their behalf. The optimal contract enables monitoring at a level privately optimal for FCs absent risk sharing motives (determined by their initial endowment ω), while simultaneously enabling FCs to trade to their privately optimal level of risk sharing absent monitoring motives (taking into account their collective market power).

How is this possible when (i) trading and monitoring are noncontractible and (ii) FMs cannot commit to any strategy? The answer is that the contract works *indirectly*, in a manner reminiscent of mechanism design. The contract allocates claims on the fund's final value (via

ϕ^*), allows for upfront transfers (via f^*), and calibrates the risk tolerance of the decision-makers (via τ^*). These parameters are chosen so that FMs, acting purely in their own interest, voluntarily trade to a point where the fund's final holdings result in monitoring and ownership levels that are optimal from the FCs' perspective. The contract *effectively* implements the FCs' full-commitment optimum without assuming any commitment.

The role of the fee. The fee f^* serves two functions. First, it compensates FMs for their monitoring costs: the FMs directly bear the cost $c(m^C)$, and without compensation they would defect from the fund and enjoy monitoring benefits for free. Second, it compensates FMs for selling part of their endowment to FCs. When FMs join the fund, each receives a pre-trade endowment of $\frac{\phi^*}{\tau^*}(\omega + \tau^* \frac{1-\omega}{1-\lambda})$, which is smaller than their initial endowment of $\frac{1-\omega}{1-\lambda}$. Conversely, FCs start with an endowment in the fund that is *higher* than their initial endowment, reflecting a reallocation of some FM endowment to FCs. The fee charges FCs for this purchase.

Remark 3. The fee f^* compensates FMs for their expected equilibrium monitoring costs as well as for the value of any endowment they sell to FCs when joining the fund.

Positive fund formation cost. We have so far conducted the analysis as if $\epsilon = 0$. Reintroducing the small fund formation cost $\epsilon > 0$ leaves the structure of the optimal contract unchanged but reduces surplus mechanically by ϵ. Since the cost is borne at the fund-formation stage and FMs were already being made exactly indifferent between joining and not joining the fund, whereas the FCs were obtaining a positive payoff Π^C_{FC}, the FCs must bear the cost either directly or (if FMs expend the cost) by paying a commensurately higher fee, resulting in a reduction in the aggregate FC payoff at the optimum from Π^C_{FC} to $\Pi^C_{FC} - \epsilon$. Since ϵ is small, the qualitative features of the optimal hedge fund contract—positive monitoring, substantial managerial co-investment, and a non-trivial mass of managers—are preserved.

Delegation as a commitment device. An important insight emerges from comparing the fund's trading behavior to that of a proprietary blockholder. In the APZ model (to be discussed fully in Section 4), a proprietary blockholder cannot stop trading short of the competitive allocation—an endowment effect means that from any position below the risk-sharing optimum, there are always marginal gains from buying a bit more. This erodes the blockholder's market power entirely, so that she is neither able to take price impact into account or limit her own ex post monitoring.

In our model, the delegation contract effectively enables commitment to a trading strategy that exploits the fund's collective market power. The contract does not directly tell managers to stop trading—rather, it gives them incentives (through τ^* and ϕ^*) that cause them to stop at a position where the FCs' effective holdings reflect their market power. In this sense, delegation serves as a commitment device for the FCs' trading strategy, even though no individual agent commits to a trading or monitoring strategy and the contract also cannot specify trading and monitoring directly.

The idea that delegation contracts can induce a degree of valuable commitment is well established in the literature in other settings. For example, a number of papers show that firm owners can benefit from delegating decisions to managers via optimal compensation contracts to commit to strategic actions in product markets (see, e.g., Fershtman and Judd, 1987, Sklivas, 1987, Dewatripont, 1988, and Fershtman, Judd, and Kalai, 1991).

2.3.5 The optimal fund and its real-world counterpart: activist hedge funds

We now argue that the optimal contract derived above maps remarkably well onto the institutional reality of hedge funds, with hedge-fund clients represented by FCs whose endowment ω of the risky asset is significant. The hedge fund-like representation is valid for any $\omega \in (0, \hat{\omega})$, but more relevant for larger ω in that range.

In the real world, hedge fund clienteles are overwhelmingly drawn from wealthy individuals and institutions: minimum net-worth requirements, together with the "qualified purchaser" thresholds embedded in the U.S. regulatory framework, effectively constrain access to high-net-worth investors. At the same time, wealth and exposure to risky assets are tightly linked. Based on the 2019 and 2022 Surveys of Consumer Finances, the Federal Reserve Board of

Governors (2023) estimates that stock-market participation rises steeply with income: 34% (for the bottom half of the income distribution), 78% (for the 50–90th percentile), and 95% (for the top 10%). The pre-delegation endowment ω carried by a hedge fund client is therefore plausibly high. Finally, hedge fund clients are sophisticated investors who may have access to gatekeepers (e.g. investment consultants, wealth management offices of private banks) who shop among competing managers and negotiate the terms of their participation, which is exactly the bargaining power assumption underlying the analysis of this subsection.

What does our model say about a fund formed for such high-ω FCs? The optimal contract requires a non-trivial mass of managers (τ^* increases in ω), substantial managerial co-investment (ϕ^* increases in ω), and meaningful monitoring, since the FMs' effective stake equals ω itself. Each of these predictions matches documented features of hedge funds. Hedge fund managers self-invest substantially in their own funds: estimates range from 7% of AUM in Agarwal, Daniel, and Naik (2009) to 20% in He and Krishnamurthy (2013). Further, hedge funds are well known for playing a far more active role in monitoring portfolio firms than other classes of intermediary (Brav, Jiang, and Kim 2010). In short, when our model is evaluated at parameters that empirically correspond to hedge-fund clienteles—significant ω, FCs with bargaining power—the resulting optimal fund matches the salient institutional features of hedge funds.

2.4 Mutual fund-like outcomes: two paths to the same allocation

The optimal contract characterized in Proposition 1 and its hedge-fund interpretation account for one important slice of the asset-management industry, but they leave a significant part of the industry unexplained. Mutual funds, which are the largest class of asset managers by both number of clients and assets under management, look very different from the activist hedge fund of Section 2.3. Mutual fund managers typically have essentially no skin in the game, do little or no firm-level monitoring, and serve clients whose primary benefit from joining the fund is diversification rather than governance. Here we show that our framework is able to deliver such fund structures as well, in two distinct ways.

As in Section 2.3, we proceed for expositional clarity by first carrying out the analysis *as if* the fund formation cost ϵ were zero, and then add it back in to show how it modifies the contracting outcomes.

Each of the two paths to mutual funds starts from the same model, but invokes a different economic force, and each delivers the same *mutual fund-like* allocation: a fund with a vanishingly small mass of managers ($\tau \to 0$) who have vanishingly small skin in the game ($\phi \to 0$); zero monitoring ($m \to 0$); and an FC stake equal to $\alpha^C = \frac{\lambda(1+\omega)}{1+\lambda}$, the same risk-sharing allocation that the FC-optimal contract delivers in the hedge-fund case. The two paths differ only in *why* monitoring vanishes and in *who* captures the resulting risk-sharing surplus, not in the allocation itself. The mutual fund pattern is therefore a robust prediction of the model, since each of the two forces in isolation is enough to produce it, and both are simultaneously plausible features of the real mutual fund industry.

2.4.1 Path 1: financial experts have bargaining power

The first way to deliver a mutual fund-like outcome is to reverse the bargaining power assumption that underpins our hedge fund derivation. Suppose that, instead of the contract being designed to maximize FC payoffs, it is designed to maximize FMs' aggregate profit subject to the FCs' participation constraint. We model this as follows. A monopolistic fund makes a take-it-or-leave-it offer to the nonfinancial investors. We assume that the offered contract maximizes the rents that can be extracted from nonfinancial investors, subject to their willingness to join the fund. Such a fund must also ensure the participation of an optimally chosen subset of financial investors who will act as fund managers.[8] This case is the natural one when asset management firms have market power vis-à-vis a fragmented retail clientele.

Proposition 2. *For any $\omega < \lambda$, a profit-maximizing fund with $\tau \to 0$ and $\phi \to 0$ can be formed that undertakes no monitoring and gives fund clients the same degree of risk sharing*

[8] Even though only a subset of financial investors act as FMs, we can think of the monopolistic fund as maximizing the payoff of all financial investors if the fund's profits are shared among the entire population of financial investors.

as the FC-optimal contract characterized in Proposition 1.

The intuition behind this result can be summarized as follows. FCs cannot obtain risk sharing without joining the fund, so they must pay for any incremental risk-sharing benefit. However, FCs *can* enjoy the benefits of monitoring without joining the fund (by free-riding on monitoring undertaken on behalf of the fund's own holdings), so they will not pay fees for monitoring. Since any monitoring the fund offers must be compensated through fees in order to satisfy the FMs' participation constraint, offering monitoring strictly reduces fund profits. The profit-maximizing fund therefore offers no monitoring at all. Once monitoring is dropped, the fund's remaining role is purely to execute risk-sharing trades on behalf of FCs. With no monitoring to incentivize, there is no reason to load FMs with co-investment, so the optimal skin-in-the-game parameter satisfies $\phi^* \to 0$. Because the FMs serve only as a trading vehicle, the optimal mass of managers shrinks to zero as well: $\tau^* \to 0$. On the other hand risk-sharing benefits *are* excludable since an FC who declines to join the fund cannot secure better risk sharing on her own. The fund can therefore extract fees from FCs in exchange for risk-sharing services up to FCs' aggregate willingness to pay, the optimal level of which corresponds exactly to delivering FCs the risk-sharing allocation $\alpha^C = \frac{\lambda(1+\omega)}{1+\lambda}$ derived in Proposition 1. The full surplus from this risk-sharing service accrues to FMs as profits. The resulting allocation is exactly the mutual fund-like one described above: no monitoring, vanishing managerial skin in the game, vanishing measure of managers, and FC stake equal to $\frac{\lambda(1+\omega)}{1+\lambda}$.

Adding back the fund formation cost $\epsilon > 0$ again leaves the structure of the contract unchanged. Since FMs design the contract to maximize their own profits, they continue to extract the full risk-sharing surplus from FCs via high fees, but must pay the cost of forming the fund. Thus their aggregate profit is reduced by ϵ.

2.4.2 Path 2: small client endowments ($\omega \to 0$)

A second way to obtain a mutual fund-like outcome is to evaluate the FC-optimal contract of Section 2.3 itself at parameters that correspond to clients with vanishingly small risky asset endowments. Recall that in the FC-optimal contract the FMs' effective stake is exactly ω, the FCs' initial endowment, because that is the level at which monitoring is privately optimal for the FCs. As $\omega \to 0$, the FC-optimal contract therefore requires $\tau^* \to 0$ and $\phi^* \to 0$: the fund employs a vanishing mass of managers with vanishing skin in the game, who undertake vanishing monitoring. The FC stake collapses to $\alpha^C \to \frac{\lambda}{1+\lambda}$, their optimal risk-sharing stake evaluated at $\omega = 0$, and the residual surplus accrues to the FCs. The contract resembles a vehicle whose only function is to deliver the FCs' privately preferred level of risk sharing.

The point of Path 2 is that even without modifying bargaining power (Path 1), the FC-optimal contract itself comes to resemble a mutual fund-like vehicle as client endowments shrink. These are the cases where no monitoring is demanded and none is provided. A potential real world interpretation of this pathway is via a commoditized, competitive fund sector serving less wealthy clients. Both forces, managerial bargaining power and small client endowments, push the optimal contract toward the same allocation.

Reintroducing the small fund formation cost $\epsilon > 0$ leaves this conclusion qualitatively unchanged. As in Section 2.3, the aggregate FC payoff is reduced by ϵ along Path 2, while the structure of the contract is unchanged, though with lower fees than under Path 1.

2.4.3 The optimal fund and its real-world counterpart: mutual funds

The model's predictions in these regimes align closely with documented features of mutual funds. Khorana, Servaes, and Wedge (2007) document that some 57% of mutual fund managers do not invest at all in their constituent funds, and the average self-investment among the rest is 0.04%, consistent with our prediction that $\phi \to 0$. Mutual funds are also well known for being muted in their engagement of portfolio firms (e.g., Bebchuk, Cohen, and Hirst 2017), consistent with our prediction of zero monitoring.

A further prediction of our model is that the optimal mutual fund operates with a vanishingly small mass of managers relative to the optimal hedge fund, since $\tau^* \to 0$ in both paths, as opposed to a non-trivial τ^* in the hedge-fund case. The number of investors per manager should therefore be significantly larger in mutual funds than in hedge funds. The institutional evidence is consistent with this prediction. The Investment Company Institute's 2024 Investment Company Fact Book reports that, at year end 2023, U.S. registered mutual funds collectively served roughly 116 million shareholders drawn from 68.7 million households. With around nine thousand U.S. mutual funds and typically only a small portfolio management team per fund, the average mutual fund manager is responsible for tens of thousands of investors. Hedge funds, by contrast, are legally restricted to a small clientele as they are typically organized to avoid the requirements of the Investment Companies Act by relying on clause 3(c)(1), which restricts them to no more than 100 investors, or clause 3(c)(7) that restricts them to only qualified purchasers. Combined with comparable or larger investment teams per fund, this regulatory ceiling implies investor to manager ratios for the typical hedge fund that are much smaller than for the typical mutual fund. The model thus captures not only the differences in skin in the game and monitoring between the two fund types, but also the asymmetry in the size of their managerial workforce relative to the size of their clientele.

2.5 When do we get hedge funds versus mutual funds?

The analysis of Sections 2.3 and 2.4 admits a clean two-dimensional summary along two primitive factors that the model isolates: client endowments (a proxy for client wealth and direct exposure to the risky asset), and the bargaining power of fund clients vis-à-vis fund managers (which, as stated above, is the way we refer to surplus extraction ability). Hedge fund-like outcomes with positive monitoring, substantial managerial co-investment, and a non-trivial mass of managers, arise only in the corner where *both* factors take high values: clients with significant endowments who have high bargaining power. In every other configuration the model predicts a mutual fund-like outcome, albeit for different economic reasons.

This taxonomy is illustrated in Table 1. When clients have low endowments and low bargaining power, the resulting fund is a mutual fund both because clients with little exposure to the risky asset would not privately demand monitoring (Path 2), and because managers with bargaining power would not voluntarily supply it (Path 1). In this case, the mutual fund offers a significant amount of risk sharing for its clients, but extracts full surplus, therefore charging high fees. When endowments are low but client bargaining power is high, i.e., the case of a fragmented fund sector serving less wealthy clients, the contract still collapses to a mutual fund because the FC-optimal level of monitoring vanishes as $\omega \to 0$ (Path 2). In this case, access to mutual funds affords significant risk sharing benefits to less wealthy clients, who additionally don't need to pay much for it because the fund management sector is fragmented or commoditized, and hence fees are low. When endowments are high but client bargaining power is low—the case of wealthy nonfinancial investors who do not have access to effective gate keepers, perhaps because the asset management industry is highly concentrated—managers strip monitoring out of the contract to maximize profits (Path 1), again delivering a mutual fund. Compared to the case of less wealthy clients, risk sharing gains are lower in this case, but clients must pay in full for them, resulting in "medium fee" mutual funds. Only when endowments are high *and* client bargaining power is high does the FC-optimal contract emerge as derived in Section 2.3, yielding the activist hedge fund. Such hedge funds charge fees not just for risk sharing but also for monitoring services, i.e., for their activism.

	Low FC bargaining power	**High FC bargaining power**
Low ω	**Mutual fund** with high fees (less wealthy clients with big mutual fund companies)	**Mutual fund** with low fees (fragmented asset management industry serving less wealthy clients)
High ω	**Mutual fund** with medium fees (wealthy clients without gate keepers; monitoring stripped out for profit)	**Hedge fund** (wealthy clients with gate keepers and a desire for high monitoring)

Table 1: When do we get hedge funds versus mutual funds? The fund type that emerges as a function of client endowments ω and the bargaining power of fund clients vis-à-vis fund managers.

This taxonomy underscores that hedge fund-like governance is a *fragile* outcome, since it requires the simultaneous presence of two specific conditions and breaks down whenever either is absent. Conversely, the mutual fund-like outcome is overdetermined, arising whenever *either* condition fails. This endogenous asymmetry is consistent with the empirical fact that mutual funds outweigh hedge funds in both number and aggregate assets under management, and that activism is limited to a comparatively small slice of the asset management industry.

2.6 Additional applied implications

Block size and monitoring intensity. Our results imply that block size may be a poor predictor of monitoring intensity. Under proprietary blockholding, larger stakes imply more monitoring because stake size directly determines monitoring intensity. Under delegated blockholding, the fund's internal incentive structure separates these two forces. For example, the total delegated block size for our hedge fund is $\omega + \frac{\lambda}{1+\lambda}(1+\omega)$, which is increasing in both the number of FCs (λ) and the aggregate FC endowment (ω). Monitoring intensity, however, depends on the FMs' skin in the game, since incentives track the monitor's own exposure, not the vehicle's size, whenever ownership and control are separated. This skin in the game, in turn, depends only on ω. Hedge funds with many investors holding limited endowments (large λ, small ω) can have large blocks with limited monitoring, while hedge funds with fewer investors but higher endowments (small λ, relatively large ω) can hold smaller blocks but monitor significantly more. In addition, mutual funds do not monitor at all, but can hold very large blocks when the nonfinancial investor population (λ) is large. Nockher (2022) provides supporting evidence that smaller blockholders tend to be more engaged monitors.

The need for index funds in governance. At the broadest level, our analysis *indirectly* highlights a role for index funds in corporate governance. Our model speaks to the concentrated holding choices of *active* funds who make deliberate portfolio decisions. We find that active funds *choose* to hold smaller positions than their full risk-bearing capacity would imply, and expend suboptimally low monitoring resources. These findings must be viewed in the context of the emergence of passive (index) funds which, purely by virtue of their size, *mechanically* hold concentrated positions. If active funds limit their holdings and monitoring, as our results suggest, it becomes all the more important to understand the role of index funds in governance (see Brav, Malenko, and Malenko 2022, and Corum, Malenko, and Malenko 2023).

3 The complementarity of investor types

Since we assume that FMs can coordinate on monitoring *within* the fund, it is natural to wonder whether financial investors could form a fund among themselves that monitors without involving FCs. The answer is no.

Proposition 3. *FM-only funds that monitor cannot exist.*

In any fund composed solely of financial investors, some subset would have to monitor and thus bear costs. But any investor in that subset can leave the fund, trade independently, and enjoy the same cash-flow payoffs without paying monitoring costs. To retain them, the non-monitoring members would have to compensate them. But this means monitoring costs are shared among all members, and the argument applies again: any non-monitoring member would prefer to leave. The only members who would not prefer to leave are those paying no monitoring costs. However, if no one pays, no monitoring occurs.

The presence of *other* investors willing to compensate FMs for monitoring is therefore essential. Nonfinancial investors fulfill this role precisely because they value the *excludable* risk-sharing benefits of fund participation enough to justify bearing the monitoring costs.

Remark 4. The endogenous formation of monitoring vehicles in financial markets requires two complementary groups: one that can provide an excludable service (risk sharing) making it worthwhile for the other to pay for a public good (monitoring).

This principle has a natural real-world interpretation. Asset management firms bundle risk-sharing services with governance services. The bundling is essential: without the risk-sharing motive that draws clients into funds, there would be no vehicle through which monitoring costs could be financed.

It is important to note, however, that the complementarity of investor types is *necessary* but not *sufficient* for monitoring. The presence of both nonfinancial and financial investors within a single vehicle is required for any monitoring to occur, but it does not by itself guarantee that monitoring will occur. As we showed in Section 2.4, mutual fund-like vehicles bring together both investor types but undertake no monitoring at all.

4 Proprietary blocks: The APZ benchmark and comparison

Having established our main results on delegated blockholding, we now briefly summarize the proprietary blocks benchmark of Admati, Pfleiderer, and Zechner (1994) and provide a detailed comparison of how and why our results differ from theirs.

4.1 The proprietary blocks benchmark

Consider the same financial market with a single firm and a unit continuum of CARA investors with risk tolerance ρ. To make blockholding proprietary, assume that a measure $\lambda < \frac{1}{2}$ of these investors are *exogenously* aggregated into a single entity, L, who trades strategically and can monitor the firm. The remaining $1 - \lambda$ investors act competitively. L has risk tolerance $\rho\lambda$ (matching the aggregate risk tolerance of the nonfinancial investors in our delegated model) and an endowment of $\omega \in (0, \lambda)$ shares, while the remaining investors have aggregate endowment $1 - \omega$. The timing is identical: trading at date 1, monitoring at date 2, and payoffs at date 3. Crucially, L cannot commit to a final round of trade at date 1 or to a monitoring level at date 2. This is (a slightly simplified version of) the classic APZ model. We state their main result below:

Proposition 4. (Admati, Pfleiderer, and Zechner 1994, Proposition 4) *As long as $\Psi(\cdot)$ is strictly concave, there exists a unique globally stable allocation $\alpha_G = \lambda$, which coincides with the competitive equilibrium allocation.*

The key implication is that the possibility of monitoring does not affect the degree of risk sharing in equilibrium. The inability to commit to a final round of trade erodes L's strategic advantage: an endowment effect induces L to trade all the way to the risk-sharing optimum. From any position strictly below λ, L is tempted to buy a bit more because the current holding is now part of her endowment, and from this endowment there are always incremental risk-sharing gains despite having to pay the full value of future monitoring.

4.2 Comparison: Delegated versus proprietary blocks

We are now in a position to provide a detailed comparison between the proprietary and delegated blocks cases. APZ find that the exogenously large proprietary trader acquires the competitive allocation and monitors at that level. Under optimal delegation, the picture changes in two fundamental ways, both driven by the separation principle.

Less risk sharing. In either configuration, our optimal fund provides less risk sharing. We discuss the case of hedge funds and mutual funds in turn.

Hedge funds. Our optimal hedge fund holds less of the risky asset than what perfect risk sharing (and thus the APZ model) would imply. Specifically, the fund's total holding is $\omega + \frac{\lambda(1+\omega)}{1+\lambda}$, which is less than the competitive risk-sharing allocation of $\lambda + \tau^*$ for an investor group of the same size. The difference arises because the delegation contract effectively enables commitment to a trading strategy that exploits market power. In the APZ model, there is no vehicle for such commitment, so the large trader cannot limit her trading to take price impact into account. In our model, the contract parameters (τ^*, ϕ^*) shape the FMs' incentives so that they voluntarily stop at a position reflecting the FCs' collective market power, without any party actually committing to anything.

Mutual funds. The case of mutual funds is simpler. A mutual fund offers the same degree of risk sharing to the FCs as the optimal hedge fund, but optimally attracts a "measure zero" of FMs (either because $\omega \to 0$ or because FMs have bargaining power). Thus, the fund holds a block of size $\frac{1+\omega}{1+\lambda}\lambda$, i.e., less than the competitive risk-sharing allocation of λ for an investor group of the same size. The intuition is the same: the contract enables the exploitation of market power.

Corollary 1. *Regardless of whether a hedge fund or mutual fund is formed, it holds strictly less of the risky asset than a proprietary blockholder with the same risk tolerance would hold.*

Less monitoring. When a hedge fund is formed, monitoring under delegation occurs at a level corresponding to the FMs' effective stake ω, which is strictly less than the fund's total holding and strictly less than the APZ blockholder's stake of λ. This is a direct consequence of the separation principle: delegation decouples monitoring from overall ownership by assigning monitoring to a subset of fund participants (the FMs) whose effective stake is calibrated to the FCs' initial endowment. The comparison for mutual funds is simpler, since the optimal mutual fund undertakes no monitoring whatsoever (either because $\omega \to 0$ or because FMs have bargaining power). The next result follows in a straightforward manner from these observations, and hence is stated without proof.

Corollary 2. *Regardless of whether a hedge fund or mutual fund is formed, the fund undertakes strictly less monitoring than a proprietary blockholder would if they held a block of identical size.*

4.3 Welfare comparisons

We compare four settings: the first-best social planner's optimum (FB), the APZ proprietary blocks benchmark (APZ), our delegated blocks equilibrium with hedge funds (DB_{HF}) and with mutual funds (DB_{MF}), and autarky (A), where nonfinancial investors hold their endowment and no monitoring occurs.

In each setting, all financial investors equally share the residual risk not held by nonfinancial investors. We can therefore index welfare by two variables: the effective stake of the monitoring entity, α_M^S, and the effective stake of the λ-sized group (either L or FCs), α_R^S, where $S \in \{FB, APZ, DB_{HF}, DB_{MF}, A\}$.

The first best has $\alpha_M^{FB} = 1$ and $\alpha_R^{FB} = \lambda$: monitoring and risk sharing are both optimal. The APZ benchmark has $\alpha_M^{APZ} = \alpha_R^{APZ} = \lambda$: risk sharing is optimal but monitoring is suboptimally low (the monitoring entity is the same as the risk-sharing entity). Our delegated blocks equilibrium under hedge funds has $\alpha_M^{DB_{HF}} = \omega < \lambda$ and $\alpha_R^{DB_{HF}} = \alpha^C = \frac{\lambda(1+\omega)}{1+\lambda} < \lambda$: both monitoring and risk sharing are below the APZ benchmark. Our delegated blocks equilibrium under mutual funds has $\alpha_M^{DB_{MF}} = 0 < \alpha_M^{DB_{HF}}$ and $\alpha_R^{DB_{MF}} = \frac{\lambda(1+\omega)}{1+\lambda} = \alpha_R^{DB_{HF}}$:

risk sharing is identical to the case of delegated blocks with hedge funds but monitoring is lower. Autarky has $\alpha_M^A = 0$ and $\alpha_R^A = \omega$: no monitoring and the worst risk sharing.

This gives a clear welfare ranking:

$$W^{FB} > W^{APZ} > W^{DB_{HF}} > W^{DB_{MF}} > W^A.$$

The APZ benchmark is inferior to the first best because of suboptimal monitoring. Our delegated blocks equilibrium, under either configuration, is inferior to APZ because of both lower monitoring and inferior risk sharing. However, delegated blockholding does provide welfare improvements relative to autarky. With hedge funds, both monitoring and risk sharing are higher than under autarky, whereas for mutual funds, only risk sharing is higher.

Interpreting the ranking. An important caveat applies to this comparison. The APZ benchmark takes the existence of a proprietary trader with risk-bearing capacity $\rho\lambda$ as exogenously given. Our model endogenizes the formation of the block by explicitly modeling the delegation process. The welfare comparison between APZ and our model is therefore best understood as comparing two different market structures—one with exogenous proprietary blocks, one with endogenous delegated blocks—rather than as a statement that delegation "destroys value." In a world where proprietary blockholders do not exist and investors must delegate to participate in financial markets, autarky is the relevant benchmark, and our optimal funds provide clear gains relative to autarky.

4.4 The economic forces underlying the differences

In this section, we explore the underlying economic forces that drive differences between the proprietary blockholder setting of APZ and the delegated blockholder setting we model here. There are two key differences between these two settings. First, under delegated blockholding, ownership and control are separate. We refer to this as the "separation friction" in what follows. Second, under delegated blockholding, individual nonfinancial investors can choose to free ride on the fund. We refer to this as the "free riding friction" in what follows. We show that both the separation friction and the free riding friction are *individually necessary* to distinguish our outcomes from those of APZ in the following sense: If either one of them fails, there exist intuitive conditions under which the APZ result can be reinstated. In particular, we show that separation is necessary whenever risk sharing needs are substantial, while free riding is necessary whenever recontracting is relatively costless.

4.4.1 The separation friction

The question we ask here is as follows. Suppose we eliminate the separation friction, i.e., a hired fund manager does not need to be incentivized, but trades "as if" she personally has the full risk tolerance of all investors in the fund and monitors "as if" she owns the fund's full block. Would the APZ result be reinstated? Eliminating separation, by construction, ensures that monitoring occurs at the full ownership of the block. However, in the presence of free riding, it will still be necessary to convince each individual fund client to join the fund. Would free riding by itself ensure that risk sharing is suboptimal and monitoring is below the APZ level?

In the absence of separation, there would be no reason to employ more than a vanishingly small number of fund managers, so we assume $\tau \to 0$, i.e., effectively a single manager. Since this manager fully internalizes the risk tolerance of the fund's investors, if all nonfinancial investors and the fund manager were to join the fund we would recover the APZ result, i.e., the manager would trade to the same globally stable allocation as a proprietary blockholder with the same risk tolerance as the fund, and monitor at that level. If such a fund could be formed, i.e., if nonfinancial investors with full free riding ability and the fund manager would join, then we conclude that the separation friction is a necessary condition for our result that delegation leads to under-monitoring and under-risk-sharing relative to the APZ benchmark. We have the following result, which applies regardless of whether FCs or FMs have bargaining power.

Proposition 5. *There exist parameters such that, in the absence of the separation friction, equilibrium risk sharing and monitoring coincide with the APZ proprietary blocks case.*

Without separation, the fund manager will, by definition, behave like the APZ proprietary blockholder and monitor at the level implied by the fund's entire stake. In the case where FCs have bargaining power, the (single) manager can always choose not to join the fund (in which case the fund doesn't form), and so must be compensated for the full anticipated monitoring cost via fees. In the case where FMs have the bargaining power, FCs will pay the full incremental benefits from joining the fund to FMs. In either case, the ultimate question is whether the risk-sharing benefits of joining an "APZ-like" fund are sufficiently large to offset the higher monitoring costs. If not, then in the case where the FCs have bargaining power they would not be willing to pay sufficient fees to induce the FM to join the fund, while in the case where the FMs have bargaining power the FCs would not join the fund that the FM would offer them. For sufficiently small ω, the risk sharing benefits are high enough relative to the anticipated monitoring cost, and the APZ outcome is realized in the absence of the separation friction. In such cases the separation friction is necessary to deliver our result that both risk sharing and monitoring are lower than in the APZ outcome. Notably, however, since monitoring is higher in the APZ case than in our main model, it is straightforward to show that the attainment of the APZ outcome without separation will be feasible for a *smaller* range of ω than that required to form the optimal hedge fund given in Proposition 1.

4.4.2 The free riding friction

Next we ask what would happen in the absence of the free riding friction. Eliminating free riding effectively implies that nonfinancial investors make decisions as a group, i.e., they will form a fund if their joint payoff in that fund is better than their joint payoff in autarky.

In APZ, an endowment effect induces the large trader, L, to trade all the way to the risk sharing optimum. Starting from any stake less than the risk-sharing optimum, there will always be at least some incremental risk sharing gains by buying a bit more, despite having to pay the full value of future monitoring in making such purchases. In the optimal hedge fund contract derived above, FCs end up with a stake of $\frac{\lambda}{1+\lambda}(1+\omega)$, which is greater than their initial endowment of ω, but monitoring occurs at a level corresponding to an ownership of ω. In the absence of free riding, will a variant of the APZ endowment effect come into play wherein the FCs now wish to recontract, ex post, with FMs to reflect their new endowment?

The relevance of this question depends, of course, on how easy it is to recontract to form a new fund. In certain settings, for example, for hedge funds with limited numbers of GPs and LPs, it may be feasible to recontract. However, in other settings, such as mutual funds with millions of investors, it may be quite challenging to do so.[9] Without taking any view on the feasibility of recontracting, we consider here the simplest case with costless recontracting.

To examine this, consider a situation where an optimal contract from above is signed and the fund trades to the stable allocation, but then an unexpected opportunity arises *ex post* to dissolve the existing fund and start a new one prior to any monitoring (if relevant) taking place.[10] Since financial investors have risk-shared perfectly among themselves we then have a repeat of the model above starting from an aggregate FC endowment of $\alpha^C(\omega) = \frac{\lambda}{1+\lambda}(1+\omega)$ instead of ω, which may lead to a new fund that may monitor at a different level.[11] We refer to such recontracting opportunities as *ex post recontracting* opportunities. Could a sequence of ex post recontracting opportunities revive the APZ result in the absence of free riding? The answer is yes regardless of whether FCs or FMs have the bargaining power.

Proposition 6. *In the absence of free riding, any sequence of costless ex post recontracting opportunities will converge to the APZ outcome.*

The convergence towards a fund that delivers the APZ outcome arises from the same logic as in APZ: once a contract is signed and trading takes place, investors will prefer to trade to a larger stake and do more monitoring because of the endowment effect. Their holdings have increased since the first round of contracting and the costs of arriving at that level are now sunk, so as a group they would like to (or agree to) pay for improved risk sharing, and are now willing to pay for monitoring at the higher level corresponding to their

[9] The literature on renegotiation with creditors starting with Gertner and Scharfstein (1991) and Bolton and Scharfstein (1996) has argued that renegotiation is more challenging when there is a larger number of creditors.

[10] Note that such a dissolution would necessarily involve a clawback of the part of the fee f^* that compensated the FMs for any anticipated monitoring costs.

[11] As noted in Lemma 4, FMs in the fund and financial investors outside the fund hold the same effective stakes after trading, so the new fund formation problem is isomorphic to the original problem with different endowments.

current aggregate stake. Thus, if FCs hold the bargaining power, they will propose a fund that gives them higher risk sharing (as long as they are not already at perfect risk sharing) and higher monitoring. If FMs hold the bargaining power, they will propose a fund that extracts the maximum surplus from the FCs, but this coincides with the fund that the FCs would propose, thus unifying the analysis of the two cases (since only surplus sharing will differ). Any sequence of such recontracting opportunities will ultimately converge to the APZ outcome, as the proof of Proposition 6 shows. Thus, we conclude that as long as it is inexpensive to recontract, free riding is necessary to prevent the equilibrium from collapsing to the APZ solution.

Free riding and recontracting

A natural follow-up question is whether recontracting opportunities will also lead to an APZ-like allocation in the *presence* of free riding. The answer is no, *even if* recontracting is perfectly costless.

Proposition 7. *In the presence of free-riding, costless ex post recontracting cannot deliver the APZ outcome.*

There are two relevant cases. First, consider the case where FCs have bargaining power. Other than in the limiting case where $\omega \to 0$, the optimal fund provides monitoring services, a non-excludable benefit. The existence of this non-excludable benefit tempts individual FCs to free-ride on the fund. The key insight is that, unlike repeated trading, which is always feasible, repeated contracting is constrained by free riding. As soon as FCs reach a new endowment level of $\omega > \hat{\omega}$ via some sequence of ex post recontracting opportunities, it is no longer possible to form a new optimal fund, by the same logic as in Lemma 2. . Since $\hat{\omega} < \lambda$, and thus the FCs' ownership stake in the largest optimal fund that can be formed, $\frac{\lambda}{1+\lambda}(1+\hat{\omega})$, is smaller than λ, it is not possible to revive the APZ result by repeatedly recontracting to the optimal fund.

For $\omega > \hat{\omega}$, is it feasible to recontract to some other, suboptimal, fund with higher monitoring, and thus sequentially reach the APZ solution? This too is not feasible, because any such alternative fund offers no greater risk sharing benefit to any individual FC (because the optimal fund maximizes their individual risk sharing payoff net of price impact), but offers more monitoring, and thus enhances the desire to stay out of the fund. As a result, for $\omega > \hat{\omega}$, no FC would ever agree to join any fund that increases monitoring relative to the optimal fund at ω. Thus, the APZ outcome cannot arise.

The case in which FMs have the bargaining power is simpler. In this case, the optimal fund offers no monitoring (see Section 2.4). Thus, there is no non-excludable benefit from the fund and no incentive to free ride. Thus, while recontracting opportunities will generate *larger* funds, ultimately resulting in perfect competitive risk sharing, nowhere along this sequence of funds will monitoring be offered. Hence, again, the APZ outcome cannot be replicated.

4.5 Endogenous blocks in APZ

APZ are not silent on endogenous block formation. In their Section VI they allow blocks to form in the presence of client free-riding and show that free-riding limits the size of the block that emerges. There are, however, several key differences between APZ's Section VI and our paper. In APZ the fund acts, by assumption, in the interest of its investors: ownership and control are never separated. In our setting the block is genuinely delegated, since our funds are formed via contracting between financial experts (who can trade and monitor) and nonfinancial clients (who cannot). This separation breaks the mechanical link between block size and monitoring: monitoring is governed by the managers' calibrated skin-in-the-game determined by the contract rather than by the fund's total holding. As a result, block size ceases to predict monitoring, and the same model rationalizes both activist hedge funds and non-monitoring mutual funds. Neither object exists in APZ. In effect, while APZ show how free riding can limit block size, their analysis cannot incorporate our core result, the separation principle, which is what allows us to characterize the conditions under which hedge funds and mutual funds emerge.

5 Discussion

5.1 What would happen if the fund formation cost were zero?

The analysis of Section 2 treated the fund formation cost $\epsilon > 0$ as a small perturbation. In both the hedge fund and mutual fund-like outcomes, ϵ alters payoffs by a small amount by either increasing fees (when FCs have bargaining power) or decreasing profits (when FMs have bargaining power), but leaves the structure of the optimal contract intact. It has an important *qualitative* role, however, in that it helps preserve the optimal fund structures of Sections 2.3 and 2.4 against an extreme and unrealistic form of unraveling. In this subsection we consider the counterfactual in which $\epsilon = 0$ and show how this changes the analysis.

5.1.1 The free entry of infinitesimally small competitors

Suppose that $\epsilon = 0$, so that a vanishingly small group of FMs (e.g., a single FM) can costlessly offer an alternative investment vehicle designed to attract a vanishingly small group of FCs (e.g., a single FC). Consider in particular what we will call a *purely passive vehicle* (PPV): a fund that promises an FC the same risk-sharing allocation $\alpha^C = \frac{\lambda(1+\omega)}{1+\lambda}$ as the hedge fund or mutual fund contract of Section 2, but because of its vanishingly small size undertakes no monitoring of its own. Because PPVs bear no monitoring costs, they can charge strictly lower fees than the hedge fund. Further, because PPVs do not extract surplus, they also charge strictly lower fees than the mutual funds under Path 1 (Section 2.4). PPVs charge the same fees as mutual funds under Path 2 (Section 2.4.2), but also deliver the same per investor allocation.

Faced with the choice between the hedge fund and a PPV that offers identical risk sharing at a lower fee, each individual FC strictly prefers to defect to the PPV, in an attempt to free ride on the hedge fund's monitoring. Iterating the argument, if all FCs can be diverted to PPVs that free-ride on the monitoring of others, then in equilibrium *no* vehicle undertakes any monitoring. Such competition would also attract investors away from the Path 1 mutual fund derived in Section 2.4 because PPVs could offer the same risk-sharing benefit at lower fees. Thus, the only funds that can survive PPV competition must offer no monitoring and fees so low that FCs retain all risk sharing surplus, i.e., Path 2 mutual funds (Section 2.4.2).

5.1.2 How $\epsilon > 0$ preserves the optimal funds in Sections 2.3 and 2.4

The unraveling argument above relies on the ability of an *arbitrarily small* mass of FMs to costlessly offer a competing vehicle to an *arbitrarily small* mass of FCs. A small but strictly positive fund formation cost $\epsilon > 0$ breaks this logic. If a deviating PPV must absorb the fixed cost ϵ in order to exist, then no individual FC—of measure zero—can finance such a deviation alone: her individual share of ϵ would be infinite. As a result, the hedge fund of Section 2.3 or Path 1 mutual fund of Section 2.4 survives as an equilibrium outcome because the *individual* deviation payoff for any FC is autarky.

We thus view the maintained assumption that $\epsilon > 0$ not as a knife-edge restriction but as a stand-in for a robust set of real-world frictions—regulatory registration fees, minimum operational and compliance costs, and the irreducible expense of negotiating contracts among participants—that prevent the unraveling described above and let the optimal funds of Section 2 emerge in equilibrium.

5.2 Opportunity costs of market participation

We have maintained the assumption throughout that nonfinancial investors are otherwise occupied and face a high opportunity cost of participating directly in financial markets. Here we compute the opportunity cost that makes each nonfinancial investor prefer to join the hedge fund or mutual fund rather than pay the cost and trade on their own.

Suppose a nonfinancial investor believes that all other nonfinancial investors will join the equilibrium fund. Then what is the minimum cost to trade that would keep them in the fund, given that they could enjoy the fund's monitoring (if any) for free while also benefiting from risk sharing? If they pay the opportunity cost, they effectively become a financial investor who, as derived above, will trade so as to perfectly share risk with the other financial investors outside the fund (and will not monitor). The minimum opportunity cost then corresponds to the difference between the payoff of a non-FM financial investor *who starts at a lower* *endowment* and a nonfinancial investor inside the fund. We have the following result.

Lemma 5. *The minimum opportunity cost that supports the formation of the optimal hedge fund or mutual fund is* $\left(\frac{\lambda-\omega}{1-\lambda^2}\right)^2 \frac{\sigma^2}{2\rho} + \frac{c^*}{\lambda}$,*where* c^* *is the equilibrium monitoring cost.*

A nonfinancial investor who pays the opportunity cost is able to enjoy the fund's monitoring for free, and thus saves the part of the per-FC fee that goes toward offsetting the FMs' monitoring cost, i.e., $\frac{c^*}{\lambda}$ (which will be zero in the mutual fund case). In addition, they will trade further than the fund would trade on their behalf, to the point where they share risk equally with all of the non-FM financial investors. This adds to the minimum opportunity cost because, as an infinitesimal trader, they can trade all the way to their risk sharing optimum (conditional on the existence of the fund) without any price impact.

6 Other realistic characteristics of asset management firms

Our model is deliberately parsimonious, yet it captures large swathes of the asset management industry in broad brush. However, some aspects of the real world are omitted. Hedge funds, for instance, routinely impose lockups and the compensation of their managers often includes convex, performance-based carry. Furthermore, asset managers often face flow-performance relationships. Finally, hedge funds and mutual funds coexist in the market, which we do not directly model. In this section we discuss how our model sheds light on these aspects of the real-world asset management industry.

6.1 Contractual features

A complete account of lockups and convex carry is methodologically demanding; such an analysis requires a separate, fully dynamic model. However, we argue below that our framework already suggests how both differences can be rationalized. A key observation is that in our model the hedge fund is the only vehicle that supplies *monitoring*, a costly and non-excludable benefit. Both lockups and convex carry, we argue, are instruments that can protect the production of monitoring once the static contract is embedded in a dynamic environment.

6.1.1 Lockups

Consider a dynamic version of the model in which the realized cost of monitoring is uncertain when the fund forms, so that monitoring cannot be fully prepaid through the up-front fee f. In a monitoring vehicle, i.e., a hedge fund, each client then has an incentive to exit before any excess cost, over and above the up-front fee, is incurred. The temptation is acute because, by the time engagement is undertaken, the target is known: limited partners can see which firm the general partners are about to engage, and any small individual client would prefer to redeem and capture the resulting improvement for free rather than remain and bear her share of the cost. This is conceptually similar to the client-side free-riding driving our central results, but operating dynamically through time rather than just through the initial entry decision. Left unchecked this could unravel the fund and eliminate monitoring. A lockup resolves the problem by letting clients commit *ex ante* not to redeem, so that client-optimal monitoring can be sustained. The argument has no force for a mutual fund, whose only output is (excludable) risk sharing, which is consistent with the absence of lockups for such funds. Our framework thus offers a rationale for lockups grounded in the protection of monitoring, a public good, and is distinct from the "limits to arbitrage" logic of Shleifer and Vishny (1997), in which lockups shield arbitrageurs from redemptions that would force untimely liquidation when prices move against them.

6.1.2 Convex carry

In our static contract, the fund's effective risk-bearing capacity and the managers' monitoring incentive are governed by the number of managers τ and their skin in the game ϕ, which pin down the fund's effective risk tolerance $\rho\tau/\phi$ and the managers' monitoring stake. However, in reality it may not always be possible to fine tune the risk tolerance of a fund's managers. Here we show that in such a case, convexity can be used to counteract limited risk tolerance.

Suppose the number of financial investors that can act as fund managers is limited, so that the fund cannot hire the full mass τ^* that the client-optimal contract calls for. Hiring a smaller mass $\tau^\# < \tau^*$ lowers the FM's aggregate risk tolerance $\rho\tau^\#$ and, under linear contracting, would preclude the fund from achieving the FC-optimal allocation. A convex carry offers a way to compensate by making managers behave as if they are more risk tolerant, so that fewer of them suffice to achieve the same allocation.

This is in general a difficult problem since, even in simple examples, the trading equilibrium is solvable only numerically. We therefore do not attempt to characterize the full optimal contract for the case of scarce managers, which lies beyond the scope of this paper. However, we develop a simple, illustrative, example. Working in our CARA–normal environment, we introduce a quadratic contract of the form $\phi V + \frac{\beta}{2}(V - \bar{V})^2$, where V is the realized value of the fund's assets, and $\bar{V}$ is its expected value. The linear part of the contract, ϕV, pins down the managers' ownership stake and, since monitoring is chosen prior to the realization of V, determines their monitoring incentive. The convex part, $\frac{\beta}{2}(V - \bar{V})^2$, governs the sharing of *deviations* from the fund's average value, and is best understood as a *side payment* between the managers and the clients.

We show that such a contract can reproduce the allocation of the FC-optimal contract in that we achieve the same monitoring level, m^C, and the same client ownership stake, α^C, as long as the aggregate risk tolerance of the available manager pool is not too low. In our example we set $\lambda = 0.4$, $\omega = 0.2$, $\sigma^2 = \rho = \mu_0 = 1$, $\mu(m) = \mu_0 + m$, and $c(m) = \frac{1}{2}\gamma m^2$ with $\gamma = 2$. We find that we can tolerate a reduction in the feasible mass from the unrestricted optimal $\tau^* \approx 0.18$ down to about 0.14, a $\sim$25% reduction, while achieving the same allocation. Full details are in the Internet Appendix.

It is important to note, however, that replicating monitoring and client ownership in this setting is not the same as delivering the clients their optimum. Because it is benchmarked to the expected fund value, the convex carry portion of the contract leaves the ownership and monitoring levels untouched, but is a state-contingent bet on which the managers are long while the clients are short, imposing a risk-sharing cost on them. The convex contract therefore cannot achieve the full unconstrained optimum of our main model.[12]

[12]The inability to solve for the full optimal contract is a common feature of the dynamic literature, see, e.g., Panageas and Westerfield (2009) and Lan, Wang, and Yang (2013).

6.2 Flow incentives

In the real world, asset managers are subject to flow-performance relationships that affect their expectation of future compensation. Such relationships may affect their incentives with respect to both trading and monitoring. For example, flow relationships could increase incentives to monitor, which could affect the optimality of our contracts. With respect to our hedge funds, any excess monitoring incentives could potentially be offset with lower skin in the game to restore optimality. However, with our mutual funds there is already no skin in the game, so the optimum may be more difficult to attain and managers may be left with higher incentives to monitor than desired by whoever has the bargaining power (see Section 2.4). It is important to note, however, that the effect of flow on incentives can be subtle. For example, the literature has shown that flow incentives can induce short-termism that *dampens* engagement incentives (e.g., Dasgupta and Piacentino 2015, Proposition 6). Furthermore, in a dynamic setting, managers' effective risk tolerance may be endogenously depressed by concern for future fees. In particular, managers may act with excessive risk aversion to protect their fees and guard against outflows (Panageas and Westerfield 2009, Lan, Wang, and Yang 2013). Such conservatism is costly for a hedge fund, which, unlike a mutual fund, needs its managers to hold the concentrated position the client-optimal contract requires. Notably, flow and carry are not independent—as we show above, carry can be used to combat excessive conservatism.

6.3 Coexistence of hedge funds and mutual funds

Our two-dimensional taxonomy gives us mutually exclusive conditions under which hedge funds and mutual funds might exist. In reality, of course, these two types of funds coexist. How can we view such coexistence through the lens of our model?

An important real-life distinction between hedge funds and mutual funds is in the types of assets they trade and the strategies they employ. Hedge funds utilize derivatives for trading and governance purposes, and generate trading profits while building disproportionately concentrated positions in firms they are targeting for governance changes (Brav, Jiang, and

Kim 2010). Such assets and strategies are not available to mutual funds due to a variety of regulations. Thus, while mutual funds do simultaneously exist, they do not replicate hedge fund portfolios in any meaningful sense. Furthermore, hedge funds are essentially restricted to directly serve only wealthy clients.

Given the existence of these different classes of assets, and the ability to invest in those assets being constrained to particular fund types, what our model could be said to provide is a positive theory of what hedge fund and mutual fund contracts, characteristics, and actions should be expected to look like in the real world (see our discussion in sections 2.3.5 and 2.4.3). That is, since specialized assets or strategies are limited to hedge funds, and hedge funds can only form for wealthy clients, if those clients have bargaining power our model would predict that hedge funds should maximize client surplus, have high skin in the game, and undertake significant monitoring. The simultaneous existence of mutual funds investing in different assets does not affect the viability of such funds since they are not a suitable alternative for accessing these assets/strategies. Put differently, rationalizing the existence of hedge fund-like outcomes in our model (i.e., the bottom right corner of Table 1) simultaneously with mutual fund-like outcomes requires some form of market segmentation, such that our model's full structure can be said to exist within each segment.

It is important to note that this argument does not mean free riding is irrelevant to hedge fund formation. Within the market segment that targets specialized strategies, free riding still operates because those clients can still benefit from the existence of the fund without joining it to the extent of their endowments. Thus, free riding limits when a hedge fund can be formed (Lemma 2 and Proposition 1), as well as how large the hedge fund can become and how much monitoring it can undertake (Proposition 7).

7 Conclusion

In financial markets with proprietary traders, risk sharing and monitoring are mechanically bundled: holding more equity simultaneously strengthens monitoring incentives. This paper shows that optimal delegation contracts unbundle these two functions. When non-expert investors delegate portfolio management to professional asset managers, the optimal contract *separates* monitoring incentives from overall fund ownership through three instruments—the measure of managers, their skin in the game, and fees. This separation implements either the clients' full-commitment optimum or profit maximization on behalf of fund managers despite the complete absence of commitment power or contractibility over trading and monitoring.

The separation principle has three broad implications. First, evaluated at parameters that correspond to wealthy clients with bargaining power, the optimal contract reproduces the salient features of activist hedge funds: substantial managerial co-investment, a non-trivial mass of managers, and meaningful monitoring of portfolio firms. The same model also rationalizes mutual fund-like outcomes with no monitoring, vanishing managerial skin in the game, and risk sharing as the sole client benefit, through two distinct economic forces: managerial bargaining power and small client endowments. Both forces deliver the same allocation, making mutual funds a robust prediction of the framework. Block size is accordingly a poor predictor of monitoring intensity. Second, delegated monitoring requires two complementary investor types: those who value risk sharing enough to pay for monitoring, and those whose incentives drive monitoring. This provides a novel rationale for why asset management firms exist as bundled institutions, though the complementarity is necessary, not sufficient, for monitoring to actually occur. Third, relative to proprietary blockholding, delegation leads to less monitoring and inferior risk sharing, but provides improvements in both dimensions relative to a world without intermediation.

Our paper provides the first analysis of optimal contracting in asset management incorporating both trading and monitoring incentives. We view our work as a first step in the direction of a full optimal contracting exercise for real world asset managers. Markets are transparent in our model, and thus trade occurs only for risk sharing reasons. In markets with asymmetric information, funds can also trade to generate alpha for their clients, an excludable benefit, and are subject to flow-performance relationships. Jointly modeling excludable alpha generation, non-excludable monitoring, risk-sharing, and flow-performance relationships represents a rich direction for future research.

Appendix

Proof of Lemma 1: In analyzing the full-commitment case, we assume that FCs collectively commit to a level of monitoring m which is publicly observed. Further, they commit publicly to a single round of trade. If they trade to a holding of α, they face a price of $\mu(m)-\frac{1-\alpha}{\rho(1-\lambda)}\sigma^2$, generating a payoff of

$$\alpha\mu(m)-c(m)-\frac{1}{2\rho\lambda}\alpha^2\sigma^2-(\alpha-\omega)\left(\mu(m)-\frac{1-\alpha}{\rho(1-\lambda)}\sigma^2\right).$$

Taking the partial derivative with respect to m yields $\omega\mu'(m)-c'(m)=0$, which does not depend on α. The SOC is satisfied, so the solution is $\omega=\frac{c'(m^C)}{\mu'(m^C)}$.

Taking the partial derivative with respect to α for given m and simplifying yields

$$\frac{(1+\omega)\sigma^2}{\rho(1-\lambda)}-\alpha\left(\frac{\sigma^2}{\rho\lambda}+\frac{2\sigma^2}{\rho(1-\lambda)}\right)=0,$$

with SOC satisfied. Solving gives $\alpha^C=\frac{1+\omega}{1+\lambda}\lambda$. ■

Proof of Lemma 2: In the full-commitment optimum, each individual nonfinancial investor has a payoff of

$$\frac{1}{\lambda}\left(\alpha^C\mu(m^C)-c(m^C)-\frac{1}{2\rho\lambda}(\alpha^C)^2\sigma^2-(\alpha^C-\omega)\left(\mu(m^C)-\frac{1-\alpha^C}{\rho(1-\lambda)}\sigma^2\right)\right).$$

If a single nonfinancial investor stays out of the fund and consumes their endowment, they receive

$$\frac{1}{\lambda}\left(\omega\mu(m^C)-\frac{1}{2\rho\lambda}\omega^2\sigma^2\right).$$

Subtracting the latter from the former yields a difference that is clearly negative when $\omega=\lambda$ (since $\alpha^C=\lambda$) and clearly positive when $\omega=0$ (by definition of α^C and m^C). Taking the ω-derivative of the difference and simplifying yields

$$-\frac{\sigma^2(\lambda-\omega)}{\lambda(1-\lambda^2)\rho}-c'(m^C)\frac{\partial m^C}{\partial\omega}<0,\quad\text{since }\frac{\partial m^C}{\partial\omega}>0.$$

Thus by continuity there exists exactly one $\hat{\omega} \in (0, \lambda)$ for which the difference is zero. ■

Proof of Lemma 3: We begin with condition (i) of the globally stable allocation. Combining definition (1) with the selected monitoring level $m^D(\alpha^D)$ defined by $\phi\alpha^D = \frac{c'(m^D(\alpha^D))}{\mu'(m^D(\alpha^D))}$ and the market clearing price (2), the optimization problem can be written as:

$$\max_{\alpha^D} \phi\alpha^D\mu(m^D(\phi\alpha^D)) - c(m^D(\phi\alpha^D)) - \frac{\phi^2(\alpha^D)^2\sigma^2}{2\rho\tau} - \Psi(\alpha_G^D) - \phi(\alpha^D - \alpha_G^D)\left(\mu(m^D(\phi\alpha_G^D)) - \frac{1-\alpha_G^D}{\rho(1-\lambda-\tau)}\sigma^2\right),$$

giving rise to the following first order condition:

$$\phi\mu(m^D(\phi\alpha^D)) - \frac{1}{\rho\tau}\phi^2\alpha^D\sigma^2 - \phi\left(\mu(m^D(\phi\alpha_G^D)) - \frac{1-\alpha_G^D}{\rho(1-\lambda-\tau)}\sigma^2\right) = 0.$$

Now, setting $\alpha^D = \alpha_G^D$ above and solving gives

$$\frac{1}{\rho\tau}\phi^2\alpha_G^D\sigma^2 = \phi\frac{1-\alpha_G^D}{\rho(1-\lambda-\tau)}\sigma^2, \text{i.e., } \alpha_G^D = \frac{\tau/\phi}{\tau/\phi + 1 - \lambda - \tau}.$$

Now, we turn to condition (ii) of the globally stable allocation to verify that $\Psi^D(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}) - \Psi^D(\omega) - \phi(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau} - \omega)P^D(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}) > 0$ for all $\omega \neq \frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$. This is equivalent to showing that $\omega = \frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$ is a global maximum of the function $\Psi^D(\omega) - \phi\omega P^D(\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau})$, i.e.,

$$\phi\omega\mu(m^D(\phi\omega)) - c(m^D(\phi\omega)) - \frac{1}{2\rho\tau}\omega^2\phi^2\sigma^2 - \phi\omega\left(\mu\left(m^D(\phi\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau})\right) - \frac{1}{\rho(\tau/\phi+1-\lambda-\tau)}\sigma^2\right).$$

To verify this we first note that the first order condition

$$\phi\mu(m^D(\phi\omega)) - \frac{1}{\rho\tau}\omega\phi^2\sigma^2 - \phi\left(\mu\left(m^D(\phi\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau})\right) - \frac{1}{\rho(\tau/\phi+1-\lambda-\tau)}\sigma^2\right) = 0$$

is satisfied at $\omega = \frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$. We then evaluate the second order condition at $\omega = \frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}$: $\phi\mu'(m^D(\phi\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}))m^{D'}(\phi\frac{\tau/\phi}{\tau/\phi+1-\lambda-\tau}) - \frac{\phi^2\sigma^2}{\rho\tau}$. This is strictly negative as long as $\Psi^D(\cdot)$ is strictly concave as required.■

Proof of Lemma 4: The per-FM effective allocation is $\frac{\phi\alpha_G^D}{\tau} = \frac{\phi}{\tau+\phi-\phi(\lambda+\tau)}$. Financial investors outside the fund hold an aggregate stake of $1-\alpha_G^D = \frac{\phi-\phi(\lambda+\tau)}{\tau+\phi-\phi(\lambda+\tau)}$, giving a per-investor allocation of $\frac{1}{1-\lambda-\tau}\frac{\phi-\phi(\lambda+\tau)}{\tau+\phi-\phi(\lambda+\tau)} = \frac{\phi}{\tau+\phi-\phi(\lambda+\tau)}$. ■

Proof of Proposition 1: To replicate a payoff of Π_{FC}^C for the FCs, (1) the fund, i.e., the FMs, must choose to monitor at level m^C, which they will only do if their own stake inside the fund is equal to ω units of the risky asset; and (2) the FCs must hold a final stake inside the fund of $\alpha^C = \frac{\lambda(1+\omega)}{(1+\lambda)}$ units of the risky asset. We choose ϕ and τ to achieve (1) and (2). For (1), we require that $\phi\alpha_G^D = \omega$. For (2), we require that $(1-\phi)\alpha_G^D = \frac{\lambda(1+\omega)}{(1+\lambda)}$. Plugging in the definition of α_G^D and solving these two equations for the two unknowns ϕ and τ yields ϕ^* and τ^* as given in the text of Proposition 1. From here forward, let the superscript $D*$ indicate that the associate function or variable is evaluated at τ^* and ϕ^*.

Next we determine the fee level f^* that just meets the participation constraint of individual FMs to ensure that the optimal mass τ^* will join the fund given the optimal skin in the game parameter ϕ^*. The fund's total endowment is given by $\omega + \tau^* \frac{1-\omega}{1-\lambda}$ (the FCs' endowment plus the FMs' share of the financial investors' aggregate endowment of $1-\omega$). The per-FM payoff for those who join the fund is given by

$$\frac{1}{\tau^*}\left[\Psi^{D*}(\alpha_G^{D*}) - \phi^*\left(\alpha_G^{D*} - \omega - \tau^*\frac{(1-\omega)}{1-\lambda}\right)P^{D*}(\alpha_G^{D*}) + \lambda f\right]. \tag{5}$$

The per-investor payoff for financial investors who do not join the fund is

$$\frac{1}{1-\lambda-\tau^*}\left[\Psi_U^{D*}(\alpha_G^{D*}) - \left(1-\alpha_G^{D*} - \left(1-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)P^{D*}(\alpha_G^{D*})\right],$$

where $\Psi_U^D(\alpha^D) = (1-\alpha^D)\mu(m^D(\alpha^D)) - \frac{(1-\alpha^D)^2\sigma^2}{2\rho(1-\lambda-\tau)}$ is the aggregate certainty equivalent payoff of the mass of $1-\lambda-\tau$ financial investors outside of the fund who hold an aggregate stake of $1-\alpha^D$ given that the fund holds a stake of α^D. By defecting from the fund unilaterally, any given FM who is supposed to join the fund can realize the latter payoff. Thus, their participation constraint will be met as long as f is set to make these two payoffs equivalent.

Defining

$$f^* = \frac{1}{\lambda}\left(\begin{array}{c} \frac{\tau^*}{1-\lambda-\tau^*}\left[\Psi_U^{D*}(\alpha_G^{D*}) - \left(1-\alpha_G^{D*}-\left(1-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)P^{D*}(\alpha_G^{D*})\right] \\ -\left[\Psi^{D*}(\alpha_G^{D*}) - \phi^*\left(\alpha_G^{D*}-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)P^{D*}(\alpha_G^{D*})\right] \end{array}\right) \tag{6}$$

ensures the participation of the requisite mass of FMs. Later in this proof, we show that the above expression for $f^*$is equivalent to the expression shown in the statement of Proposition 1.

To complete the proof, we now show that the above contracting terms lead to an aggregate payoff for the FCs of Π_{FC}^C. First, we show that the price of the risky asset in the delegated fund equilibrium is equivalent to the price in the FCs' full-commitment optimum. In the full-commitment optimum, the price is given by $\mu(m^C) - \frac{1-\alpha^C}{\rho(1-\lambda)}\sigma^2$. Replacing α^C with $\frac{\lambda(1+\omega)}{(1+\lambda)}$ yields a price of $\mu(m^C) - \frac{1-\lambda\omega}{\rho(1-\lambda^2)}\sigma^2$. In the delegated fund equilibrium, the price, evaluated at the optimal fund parameters, is given by $P^{D*}(\alpha_G^{D*}) = \mu(m^C) - \frac{1-\frac{\tau^*/\phi^*}{\tau^*/\phi^*+1-\lambda-\tau^*}}{\rho(1-\lambda-\tau^*)}\sigma^2$. It is straightforward to show the algebraic equivalence of these two prices using the definitions of τ^* and ϕ^*.

Since the level of monitoring in the delegated fund equilibrium is identical to that in the FCs' full-commitment optimum and the FCs' final holdings are identical, to complete our argument it suffices to show that FCs pay identical effective monitoring costs and trading costs across the two cases. With respect to the monitoring costs, note that the FMs directly pay the entirety of the actual costs in the delegated fund equilibrium while the FCs pay these costs in the full-commitment optimum. Thus, the aggregate fee paid by the FCs must compensate FMs for their monitoring costs. With respect to trading costs, the costs incurred by the FCs in their full-commitment optimum equal the equilibrium price times their aggregate trading quantity, or $P^{D*}(\alpha_G^{D*})(\alpha^C - \omega)$ (using the result above that the equilibrium price is equivalent to the full-commitment price). In the delegated fund equilibrium they directly pay trading costs equal to the price times their proportional stake in the fund times its overall trading quantity, or $P^{D*}(\alpha_G^{D*})(1-\phi^*)(\alpha_G^{D*} - \omega - \tau^*\frac{(1-\omega)}{1-\lambda})$. Since, as shown previously, $(1-\phi^*)\alpha_G^{D*} = \alpha^C$, the *savings* in the FCs' direct trading costs for the delegated fund equilibrium relative to their full-commitment equilibrium equal

$$P^{D*}(\alpha_G^{D*})\left[(\alpha^C-\omega)-(1-\phi^*)(\alpha_G^{D*}-\omega-\tau^*\frac{(1-\omega)}{1-\lambda})\right]=P^{D*}(\alpha_G^{D*})\left[(1-\phi^*)(\omega+\tau^*\frac{(1-\omega)}{1-\lambda})-\omega\right].$$

Thus, the aggregate fee must also transfer this amount from the FCs to the FMs.

To show that the equilibrium fee, f^* as defined in (6) accomplishes these requirements, first note that it is straightforward to show that $\frac{\tau^*}{1-\lambda-\tau^*}(1-\alpha_G^{D*})=\omega$, and since we know that $\phi^*\alpha_G^{D*}=\omega$ also holds, we have $\frac{\tau^*}{1-\lambda-\tau^*}\Psi_U^{D*}(\alpha_G^{D*})-\Psi^{D*}(\alpha_G^{D*})=c(m^C)$. We can therefore rewrite the fee f^* as

$$\frac{1}{\lambda}\left[c(m^C)+P^{D*}(\alpha_G^{D*})\left(\phi^*(\alpha_G^{D*}-\omega-\tau^*\frac{(1-\omega)}{1-\lambda})-\frac{\tau^*}{1-\lambda-\tau^*}\left(1-\alpha_G^{D*}-\left(1-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)\right)\right]$$

or

$$\frac{1}{\lambda}\left[c(m^C)+P^{D*}(\alpha_G^{D*})\left(\phi^*(-\omega-\tau^*\frac{(1-\omega)}{1-\lambda})-\frac{\tau^*}{1-\lambda-\tau^*}\left(-\left(1-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)\right)\right]$$

or

$$\frac{1}{\lambda}\left[c(m^C)+P^{D*}(\alpha_G^{D*})\left(\phi^*(-\omega-\tau^*\frac{(1-\omega)}{1-\lambda})-\frac{\phi^*\alpha_G^{D*}}{(1-\alpha_G^{D*})}\left(-\left(1-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)\right)\right]$$

or

$$\frac{1}{\lambda}\left[c(m^C)+P^{D*}(\alpha_G^{D*})\left((1-\phi^*)(\omega+\tau^*\frac{(1-\omega)}{1-\lambda})-\omega\right)\right],$$

as in Proposition 1. Thus, since the aggregate fee is λf^*, the FCs' obtain payoff Π_{FC}^C. Finally, note that Lemma 2 implies that all λ FCs find participation in the fund optimal as long as $\omega\leq\hat{\omega}$.■

Proof of Proposition 2: Our proof proceeds in three steps.

Step 1: We first show that a profit maximizing fund will offer no monitoring. Consider an arbitrary fund that offers FCs a holding of α and undertakes monitoring at some arbitrary level m. This involves a risk sharing cost of trading from ω to α at equilibrium prices and a monitoring cost of $c(m)$. We assume WLOG that both of these costs are charged to FCs (as otherwise it would have to be paid by the fund itself), but the amount of risk sharing and the amount of monitoring to be targeted by the fund is a choice of the fund. Since it has bargaining power, the fund can then extract the full surplus from FCs from joining the fund. The overall utility of FCs (net of the charges referred to above but gross of any additional surplus extraction by the fund) from being in the fund is:

$$\frac{1}{\lambda}\left(\alpha\mu(m)-\frac{1}{2\rho\lambda}(\alpha)^2\sigma^2-(\alpha-\omega)\left(\mu(m)-\frac{1-\alpha}{\rho(1-\lambda)}\sigma^2\right)\right)-c(m),$$

whereas their payoff from staying outside the fund (conditional on the fund's existence) is,

$$\frac{1}{\lambda}\left(\omega\mu(m)-\frac{1}{2\rho\lambda}\omega^2\sigma^2\right).$$

Thus, the maximal additional surplus that can be extracted from FCs is:

$$\frac{1}{\lambda}\left(\alpha\mu(m)-\frac{1}{2\rho\lambda}(\alpha)^2\sigma^2-(\alpha-\omega)\left(\mu(m)-\frac{1-\alpha}{\rho(1-\lambda)}\sigma^2\right)\right)-\frac{1}{\lambda}\left(\omega\mu(m)-\frac{1}{2\rho\lambda}\omega^2\sigma^2\right)-c(m),$$

i.e., upon simplification

$$\frac{1}{\lambda}\left(-\frac{1}{2\rho\lambda}(\alpha)^2\sigma^2+(\alpha-\omega)\frac{1-\alpha}{\rho(1-\lambda)}\sigma^2\right)+\frac{1}{2\rho\lambda^2}\omega^2\sigma^2-c(m),$$

which is clearly decreasing in the monitoring intensity m. Hence the fund will offer no monitoring, setting $m=0$.

Step 2: We now show that, given the absence of monitoring, the fund will offer FCs the same risk sharing level as the optimal fund of Proposition 1. Evaluated at $m=0$, the rents extractable by the fund are as follows:

$$\frac{1}{\lambda}\left(\alpha\mu(0)-\frac{1}{2\rho\lambda}(\alpha)^2\sigma^2-(\alpha-\omega)\left(\mu(0)-\frac{1-\alpha}{\rho(1-\lambda)}\sigma^2\right)\right)-\frac{1}{\lambda}\left(\omega\mu(0)-\frac{1}{2\rho\lambda}\omega^2\sigma^2\right),$$

which (by appeal to the computations in Lemma 1) is maximized at $\alpha=\frac{1+\omega}{1+\lambda}\lambda$. Since this fund undertakes no monitoring, it must give a vanishing stake to any FMs who are hired, i.e., $\phi \to 0$. If $\phi \to 0$ but the measure of FMs, $\tau \nrightarrow 0$ then the FMs' trading would become infinite. Hence, $\tau \to 0$ as well.

Step 3: The fund can now be written as the limit of a sequence of funds in our (τ, ϕ) representation as follows: For any ω and τ, set $\phi = \frac{1-\omega\lambda}{\lambda(1-\lambda)(1+\omega)}\tau$. Then $\alpha_G^D = \frac{\frac{\lambda(1-\lambda)(1+\omega)}{1-\omega\lambda}}{\frac{\lambda(1-\lambda)(1+\omega)}{1-\omega\lambda}+1-\lambda-\tau}$. Taking $\tau \to 0$ gives $\lim \alpha_G^D = \frac{1+\omega}{1+\lambda}\lambda$ as required. ■

Proof of Proposition 3: Suppose a fund of $\theta \in (0, 1-\lambda]$ financial investors is proposed with total post-trade stake α_θ and monitoring according to collective holdings. Each investor must pay part of (i) trading costs to reach α_θ and (ii) total monitoring cost $c(m(\alpha_\theta))$.

Any financial investor can leave, enjoy monitoring for free, and trade to $\frac{\alpha_\theta}{\theta}$ at $\frac{1}{\theta}$ fraction of trading costs. Any investor paying above-average trading costs benefits from leaving, so all must share equally. Given equal trading costs, any investor paying monitoring costs strictly prefers to leave (identical trading cost, zero monitoring cost). The only investors who would not prefer to leave pay no monitoring costs—but if nobody pays, no monitoring occurs, contradicting the premise. ■

Proof of Proposition 4: If α is L's total ownership upon entering date 2, her equilibrium monitoring level is given by $m(\alpha) = \arg\max_m \Psi(\alpha)$, where

$$\Psi(\alpha) = \alpha\mu(m) - c(m) - \frac{1}{2\rho\lambda}\alpha^2\sigma^2 \tag{7}$$

is L's certainty equivalent. The optimal monitoring level satisfies

$$\alpha = \frac{c'(m(\alpha))}{\mu'(m(\alpha))}. \tag{8}$$

If L's expected final ownership is α, the market clearing price is

$$P(\alpha) = \mu(m(\alpha)) - \frac{1-\alpha}{\rho(1-\lambda)}\sigma^2. \tag{9}$$

According to original APZ paper, from which definition (1) was adapted, an allocation α_G is globally stable iff (i)

$$\alpha_G \in \arg\max_{\alpha} \Psi(\alpha) - \Psi(\alpha_G) - (\alpha - \alpha_G)P(\alpha_G),$$

and (ii) for every $\omega \in [0,1]$ such that $\omega \neq \alpha_G$,

$$\Psi(\alpha_G) - \Psi(\omega) - (\alpha_G - \omega)P(\alpha_G) > 0.$$

Combining (7) with $m(\alpha)$ from (8) and $P(\alpha)$ from (9), the first order condition for condition (i) simplifies to

$$\mu(m(\alpha)) - \frac{1}{\rho\lambda}\alpha\sigma^2 - \left(\mu(m(\alpha_G)) - \frac{1-\alpha_G}{\rho(1-\lambda)}\sigma^2\right) = 0.$$

Setting $\alpha = \alpha_G$ and solving gives $\frac{1}{\rho\lambda}\alpha_G\sigma^2 = \frac{1-\alpha_G}{\rho(1-\lambda)}\sigma^2$, i.e., $\alpha_G = \lambda$.

For condition (ii), we verify that $\omega = \lambda$ is a global maximum of $\Psi(\omega) - \omega P(\lambda) = \omega\mu(m(\omega)) - c(m(\omega)) - \frac{1}{2\rho\lambda}\omega^2\sigma^2 - \omega(\mu(m(\lambda)) - \frac{\sigma^2}{\rho})$. The FOC is satisfied at $\omega = \lambda$, and the SOC $\mu'(m(\lambda))m'(\lambda) - \frac{\sigma^2}{\rho\lambda}$ is strictly negative when $\Psi(\cdot)$ is strictly concave. ■

Proof of Corollary 1: The hedge fund holds $\alpha_G^{D*} = \frac{\tau^*/\phi^*}{\tau^*/\phi^*+1-\lambda-\tau^*}$. Using the expressions from Proposition 1:

$$\frac{\tau^*}{\phi^*} = \frac{(1-\lambda)(2\lambda\omega+\lambda+\omega)}{1-\lambda\omega}.$$

We show $\tau^*/\phi^* < \lambda + \tau^*$. Assuming the contrary and simplifying leads to $\lambda(\omega - \lambda) \geq 0$, a contradiction since $\omega < \lambda$. Therefore $\alpha_G^{D*} < \frac{\lambda+\tau^*}{\lambda+\tau^*+1-\lambda-\tau^*} = \lambda + \tau^*$. In other words, the holding of the risky asset by the members of the optimal hedge fund (a measure λ of FCs and a measure τ^* of FMs) is small that their proportion of the aggregate risk tolerance in the economy.

The case of the mutual fund is simpler. The mutual fund always optimally attracts a measure zero of FMs (either because $\omega \to 0$ or FMs have bargaining power) and thus represents a fraction λ of the aggregate risk tolerance in the economy. However, such a fund holds a block of size $\frac{1+\omega}{1+\lambda}\lambda < \lambda$, and thus again features imperfect risk sharing.■

Proof of Proposition 5: We first check when it is better for each FC to join a fund without a separation friction. We start with the case in which the FCs have bargaining power. Consider a fund where the single (measure zero) FM acts "as if" her risk tolerance is λ (because she represents all FCs who have joined the fund) and monitors as if she holds the full block but must be fully compensated for it (otherwise she would not become a fund manager). We then compute the payoff of each nonfinancial investor from joining the fund. By analogy to the proof of Lemma 2 we have:

$$\frac{1}{\lambda}\left(\lambda\mu(\lambda) - c(\lambda) - \frac{\lambda\sigma^2}{2\rho} - (\lambda-\omega)\left(\mu(\lambda) - \frac{\sigma^2}{\rho}\right)\right),$$

which we then compare to her payoff from not joining the fund:

$$\frac{1}{\lambda}\left(\omega\mu(\lambda) - \frac{\omega^2\sigma^2}{2\rho\lambda}\right).$$

Each such nonfinancial investor would thus choose to join such a fund if and only if

$$\frac{\sigma^2}{2\rho\lambda}(\lambda-\omega)^2 > c(\lambda). \tag{10}$$

Now we note that the same condition will hold in the case where the FMs have the bargaining power, since the only difference in payoffs for the FCs will be an additional fee equal to their full incremental payoff relative to autarky from joining the fund above. As long as

$$\frac{\sigma^2}{2\rho} > \frac{c(\lambda)}{\lambda},$$

i.e., the asset is risky enough relative to the level of risk tolerance that the risk sharing benefits outweigh the per-investor monitoring cost, there exist a positive measure of $\omega > 0$ for which (10) will hold. Thus, for small enough endowments and a risky enough asset, this fund can be formed and we recover APZ.■

Proof of Proposition 6: Start with any arbitrary fund, in which FCs hold an arbitrary amount $\xi < \lambda$ and monitoring occurs at an intensity corresponding to a holding of $\chi <$

ξ, and assume that the financial investors have perfectly risk shared amongst themselves (in other words, the fund's holdings represent a globally stable allocation). Starting from this position, suppose that, when presented with an ex post recontracting opportunity, the recontracted fund selects holding level of $\Gamma(\xi) \equiv \frac{\lambda}{1+\lambda}(1+\xi)$ and monitoring occurring at intensity corresponding to a holding level of ξ. Since $\lambda \in \left(0, \frac{1}{2}\right)$, so that $\frac{\lambda}{1+\lambda} \in \left(0, \frac{1}{3}\right)$, Γ is a contraction mapping. By the contraction mapping theorem, regardless of the starting point, any sequence of ex post recontracting converges to the unique limit given by

$$\xi^* = \frac{\lambda}{1+\lambda}(1+\xi^*),$$

i.e., $\xi^* = \lambda$ which represents both the limiting holding level for the FCs and the holding level that determines the limiting monitoring intensity.

Now consider the two bargaining protocols in turn.

For the case in which FCs have the bargaining power, let us start from the hedge fund characterized in Proposition 1, $\xi = \frac{\lambda}{1+\lambda}(1+\omega)$ and $\chi = \omega < \xi$. Presented with an ex post recontracting opportunity, the FCs will (re-applying Proposition 1) wish to recontract to a fund that results in a holding level of $\Gamma(\xi)$ and a monitoring intensity corresponding to a holding level of ξ. Thus, any sequence of such ex post recontracting will result in a limiting fund in which holdings and monitoring both coincide with the APZ outcome.

If FMs have the bargaining power, starting from the mutual fund characterized in Proposition 2, $\xi = \frac{\lambda}{1+\lambda}(1+\omega)$ (since FCs are offered exactly the same risk sharing as in the hedge fund) and $\chi = 0 < \xi$ (since there is no monitoring). Presented with an ex post recontracting opportunity, the FM's will propose the fund that offers the FCs their highest possible surplus (i.e., the optimal hedge fund of Proposition 1 which gives the FCs their privately optimal full-commitment outcome) given that when free riding is impossible the FCs will always join this fund and pay their entire surplus to the FMs in fees. Thus, FMs will wish to propose a fund that offers FCs holdings of $\Gamma(\xi)$ and a monitoring intensity corresponding to a holding level of ξ. Thus, again, any sequence of such ex post recontracting will result in a limiting fund in which holdings and monitoring both coincide with the APZ outcome.■

Proof of Proposition 7: First, consider the case where the FCs have bargaining power. Consider any starting point where an optimal contract is in place with monitoring at intensity corresponding to $\omega < \hat{\omega} < \lambda$ and FC holdings of $\frac{\lambda}{1+\lambda}(1+\omega)$.

Case 1: $\frac{\lambda}{1+\lambda}(1+\omega) \leq \hat{\omega}$. A new optimal fund can form with higher monitoring corresponding to $\frac{\lambda}{1+\lambda}(1+\omega)$, but since $\hat{\omega} < \lambda$, monitoring remains below APZ. The new fund again belongs to Case 1 or Case 2.

Case 2: $\frac{\lambda}{1+\lambda}(1+\omega) > \hat{\omega}$. By Lemma 2, no optimal fund can form. Moreover, no suboptimal fund with *higher* monitoring can form either: at monitoring corresponding to $\omega' > \omega$, the maximal risk sharing achievable at the new endowment (by Lemma 1) is the optimal fund's allocation $\frac{\lambda}{1+\lambda}(1+\frac{\lambda}{1+\lambda}(1+\omega))$. But Lemma 2 shows this risk sharing is insufficient to induce FC participation even at the lower monitoring level of ω; a fortiori it is insufficient at higher monitoring.

Now consider the case where the FMs have bargaining power. Starting from the fund characterized in Proposition 2, which has FC holdings of $\xi = \frac{\lambda}{1+\lambda}(1+\omega)$ and monitoring at level $\chi = 0$. Since profit maximizing FMs will always maintain the non-excludable benefit at zero, but maximize (excludable) risk-sharing provision to FCs to maximize profits, the FMs will, given a recontracting opportunity, offer a contract with holdings of $\Gamma(\xi) \equiv \frac{\lambda}{1+\lambda}(1+\xi)$ and monitoring occurring at intensity corresponding to a holding level of 0. Since $\lambda \in \left(0, \frac{1}{2}\right)$, so that $\frac{\lambda}{1+\lambda} \in \left(0, \frac{1}{3}\right)$, Γ is a contraction mapping. By the contraction mapping theorem, regardless of the starting point, any sequence of ex post recontracting converges to the unique fixed point given by

$$\xi^* = \frac{\lambda}{1+\lambda}(1+\xi^*),$$

i.e., $\xi^* = \lambda$. Thus, the limit of a sequence of ex post recontracting delivers a fund that provides the same risk sharing benefits as the APZ allocation, but with no monitoring.■

Proof of Lemma 5: First consider the case of the optimal hedge fund. From above, the payoff of an FC in the fund is given by

$$\frac{1}{\lambda}\left(\alpha^C\mu(m^C) - c(m^C) - \frac{1}{2\rho\lambda}(\alpha^C)^2\sigma^2 - (\alpha^C - \omega)\left(\mu(m^C) - \frac{1-\alpha^C}{\rho(1-\lambda)}\sigma^2\right)\right),$$

whereas the payoff of a financial investor outside the fund is

$$\frac{1}{1-\lambda-\tau^*}\left[\Psi_U^{D*}(\alpha_G^{D*})-\left(1-\alpha_G^{D*}-\left(1-\omega-\tau^*\frac{(1-\omega)}{1-\lambda}\right)\right)P^{D*}(\alpha_G^{D*})\right].$$

A nonfinancial investor who pays the opportunity cost, however, starts from a smaller endowment, so its payoff outside the fund is

$$\frac{1}{1-\lambda-\tau^*}\left[\Psi_U^{D*}(\alpha_G^{D*})-\left(1-\alpha_G^{D*}-\frac{\omega}{\lambda}(1-\lambda-\tau^*)\right)P^{D*}(\alpha_G^{D*})\right].$$

Replacing $\Psi_U^{D*}(\alpha_G^{D*})$ with its definition $(1-\alpha_G^{D*})\mu(m^D(\alpha_G^{D*}))-\frac{(1-\alpha^{D*})^2\sigma^2}{2\rho(1-\lambda-\tau)}$, recognizing that $\alpha_G^{D*}=\alpha^C+\omega$ and $P^{D*}(\alpha_G^{D*})=\mu(m^C)-\frac{1-\alpha^C}{\rho(1-\lambda)}\sigma^2$, gives

$$\frac{1}{1-\lambda-\tau^*}\left[\begin{array}{c}(1-(\alpha^C+\omega))\mu(m^C)-\frac{(1-(\alpha^C+\omega))^2\sigma^2}{2\rho(1-\lambda-\tau^*)}\\ -\left(1-(\alpha^C+\omega)-\frac{\omega}{\lambda}(1-\lambda-\tau^*)\right)\left(\mu(m^C)-\frac{1-\alpha^C}{\rho(1-\lambda)}\sigma^2\right)\end{array}\right].$$

Replacing τ^* with$\frac{(1-\lambda^2)\omega}{1-\lambda\omega}$ and α^C with $\frac{\lambda(1+\omega)}{(1+\lambda)}$ and simplifying gives

$$\frac{1-\lambda\omega}{1-\lambda^2}\left(\mu(m^C)-\left(\frac{1-\lambda\omega}{1-\lambda^2}\right)\frac{\sigma^2}{2\rho}\right)-\left(\frac{1-\lambda\omega}{1-\lambda^2}-\frac{\omega}{\lambda}\right)\left(\mu(m^C)-\frac{1-\frac{\lambda(1+\omega)}{(1+\lambda)}}{\rho(1-\lambda)}\sigma^2\right).$$

Replacing α^C with $\frac{\lambda(1+\omega)}{(1+\lambda)}$ in the per-FC payoff expression above and simplifying yields

$$\frac{1+\omega}{1+\lambda}\left(\mu(m^C)-\left(\frac{1+\omega}{1+\lambda}\right)\frac{\sigma^2}{2\rho}\right)-\left(\frac{(1+\omega)}{(1+\lambda)}-\frac{\omega}{\lambda}\right)\left(\mu(m^C)-\frac{1-\frac{\lambda(1+\omega)}{(1+\lambda)}}{\rho(1-\lambda)}\sigma^2\right)-\frac{c(m^C)}{\lambda}.$$

Subtracting the latter from the former (noting that $\frac{1-\lambda\omega}{1-\lambda^2}-\frac{1+\omega}{1+\lambda}=\frac{\lambda-\omega}{1-\lambda^2}>0$) yields

$$\left(\frac{\lambda-\omega}{1-\lambda^2}\right)\left(\frac{1-\lambda\omega}{\rho(1-\lambda^2)}\sigma^2\right)-\left(\frac{\lambda-\omega}{1-\lambda^2}\right)\left(\frac{2-\lambda+\omega-2\lambda\omega}{1-\lambda^2}\right)\frac{\sigma^2}{2\rho}+\frac{c(m^C)}{\lambda},$$

or, simplifying,

$$\left(\frac{\lambda-\omega}{1-\lambda^2}\right)^2\frac{\sigma^2}{2\rho}+\frac{c(m^C)}{\lambda}.$$

Now consider the case of an FM-offered mutual fund (the case for $\omega \to 0$ is covered by the above). A financial investor's payoff in the mutual fund equilibrium, given that there will be no monitoring in equilibrium, is

$$\frac{1}{1-\lambda}\left[(1-\alpha^C)\mu(0) - \frac{(1-\alpha^C)^2\sigma^2}{2\rho(1-\lambda)} - \left(1-\alpha^C-(1-\omega)\right)\left(\mu(0) - \frac{1-\alpha^C}{\rho(1-\lambda)}\sigma^2\right)\right].$$

A nonfinancial investor who pays the cost to learn to trade on their own starts from a lower endowment, and thus receives

$$\frac{1}{1-\lambda}\left[(1-\alpha^C)\mu(0) - \frac{(1-\alpha^C)^2\sigma^2}{2\rho(1-\lambda)} - \left(1-\alpha^C-\frac{\omega}{\lambda}(1-\lambda)\right)\left(\mu(0) - \frac{1-\alpha^C}{\rho(1-\lambda)}\sigma^2\right)\right].$$

Since they are held to their autarky payoff by the FMs in the mutual fund, the minimum opportunity cost must set their payoff after paying the cost equal to their autarky payoff, which is

$$\frac{1}{\lambda}\left(\omega\mu(0) - \frac{1}{2\rho\lambda}\omega^2\sigma^2\right).$$

Replacing α^C with $\frac{\lambda(1+\omega)}{(1+\lambda)}$ in the former expression, taking the difference, and simplifying yields

$$\left(\frac{\lambda-\omega}{1-\lambda^2}\right)^2\frac{\sigma^2}{2\rho}$$

as required.■

Bibliography

Admati, A., P. Pfleiderer, and J. Zechner 1994. Large shareholder activism, risk sharing, and financial market equilibrium. Journal of Political Economy 102, 1097–1130.

Agarwal, Vikas, Naveen D. Daniel, and Narayan Y. Naik, 2009, Role of managerial incentives and discretion in hedge fund performance, Journal of Finance 64, 2221–2256.

Bebchuk, L. A., A. Cohen, and S. Hirst 2017. The agency problems of institutional investors. Journal of Economic Perspectives 31, 89–102.

Bhattacharya, S. and P. Pfleiderer 1985. Delegated Portfolio Management. Journal of Economic Theory 36, 1–25.

Bolton, P. and H. Rosenthal 2019. The End of the "Modern Corporation": Deregulation and Ownership of Electric Utilities. Working paper, Columbia University.

Bolton, P. and D. Scharfstein 1996. Optimal Debt Structure and the Number of Creditors. Journal of Political Economy 104, 1–25.

Bolton, P. and E. von Thadden 1998. Blocks, Liquidity, and Corporate Control. Journal of Finance 53, 1–25.

Brav, A., A. Dasgupta, and R. Mathews 2022. Wolf Pack Activism. Management Science 68, 5557–5568.

Brav, A., W. Jiang, and H. Kim 2010. Hedge fund activism: A review. Foundations and Trends in Finance 4, 185–246.

Brav, A., A. Malenko, and N. Malenko 2022. Corporate governance implications of the growth in indexing. ECGI working paper 849/2022.

Chevalier, J. and G. Ellison 1997. Risk Taking by Mutual Funds as a Response to Incentives. Journal of Political Economy 105, 1167–1200.

Corum, A., A. Malenko, and N. Malenko 2023. Corporate governance in the presence of active and passive delegated investment. ECGI working paper 695/2020.

Dasgupta A., V. Fos, Z. Sautner 2021. Institutional Investors and Corporate Governance. Foundations and Trends in Finance 12, 276–394.

Dasgupta, A., and E. Maug 2023. Delegation chains. ECGI working paper 798/2021.

Dasgupta, A. and G. Piacentino 2015. The wall street walk when blockholders compete for flows. Journal of Finance 70, 2853–2896.

DeMarzo, P. and B. Urosevic 2006. Ownership Dynamics and Asset Pricing with a Large Shareholder. Journal of Political Economy 114, 774–815.

Dewatripont, M. 1988. Commitment Through Renegotiation-Proof Contracts with Third Parties. Review of Economic Studies 55, 377–389.

Diamond, D. and P. Dybvig 1983. Bank Runs, Deposit Insurance, and Liquidity. Journal of Political Economy 91, 401–419.

Edmans, A. and C. Holderness 2017. Blockholders: A survey of theory and evidence. In: Handbook of the Economics of Corporate Governance, Vol. 1, 541–636.

Federal Reserve Board of Governors 2023. Changes in U.S. Family Finances from 2019 to 2022. Board of Governors of the Federal Reserve System Research & Analysis, October 2023.

Fershtman, C. & K. Judd 1987. Equilibrium Incentives in Oligopoly. American Economic Review 77, 927–940.

Fershtman, C., K. Judd & E. Kalai 1991. Observable Contracts: Strategic Delegation and Cooperation. International Economic Review 32, 551–559.

Gertner, R. & D. Scharfstein 1991. A Theory of Workouts and the Effects of Reorganization Law. Journal of Finance 46, 1189–1222.

Grossman, Sanford J., and Oliver D. Hart 1980. Takeover bids, the free-rider problem, and the theory of the corporation. Bell Journal of Economics 11, 42–64.

He, Z. and A. Krishnamurthy 2013. Intermediary asset pricing. American Economic Review 103, 732–770.

ICI 2024. Investment company factbook. Investment Company Institute, https://www.icifactbook.org/.

Khorana, A., H. Servaes, and L. Wedge 2007. Portfolio manager ownership and fund performance. Journal of Financial Economics 85, 179–204.

Kyle, A. and J.-L. Vila 1991. Noise trading and takeovers. RAND Journal of Economics 22, 54–71.

Marinovic, I. and F. Varas 2021. Strategic trading and blockholder dynamics. Working paper, Stanford University.

Nockher, F. 2022. The Value of Undiversified Shareholder Engagement. Working paper, University of Pennsylvania.

Piacentino, G. 2019. Venture capital and capital allocation. Journal of Finance 74, 1261–1314.

Shleifer, A. and R. Vishny 1986. Large shareholders and corporate control. Journal of Political Economy 94, 461–488.

Shleifer, A. and R. Vishny, 1997. The Limits of Arbitrage. Journal of Finance 52, 35–55.

Sklivas, S. 1987. The Strategic Choice of Managerial Incentives. RAND Journal of Economics 18, 452–458.

Internet Appendix

1 Convex carry

The client-optimal contract characterized in the main text employs a mass τ^* of fund managers under a linear sharing rule. Two features of practice motivate revisiting the number of managers. First, the supply of suitable managers may be limited, so a fund may be unable to hire the full mass τ^*. Second, hedge-fund managers are typically compensated through convex, performance-based carry. This appendix asks whether a convex carry can substitute for managerial mass: can a fund that hires a smaller mass $\tau^\# < \tau^*$ of managers, but pays them a convex carry, still reproduce the client-optimal outcome?

With a quadratic carry, we show that we can reproduce the same allocation—the same monitoring level m^C and the same client ownership stake α^C—but only for a mass of managers above a feasibility floor. The convexity raises the managers' willingness to hold the concentrated position that supports monitoring at the client-optimal level, so fewer of them suffice. However, it does so by introducing an appetite to over-accumulate that, beyond a point, erodes global stability, as we discuss below. Furthermore, while we can achieve the same ownership allocation, the convexity that is paid to the managers is borne by the fund clients, imposing a cost on them, thus not achieving the full unconstrained optimum of our main model.

1.1 The quadratic carry contract

We reuse our baseline model. Recall that the FC-optimal allocation that generated the hedge fund was characterized by a fund holding of $\alpha_G^D = \frac{\tau^*/\phi^*}{\tau^*/\phi^*+1-\lambda-\tau}$, of which FCs hold $\alpha^C = \frac{\lambda(1+\omega)}{1+\lambda}$ and FMs hold $\phi^*\alpha_G^D = \omega$, where$\tau^* = \frac{(1-\lambda^2)\omega}{(1-\lambda\omega)}$.

Now, assume the FCs are constrained to hire a smaller mass $\tau^\# < \tau^*$ of managers, with aggregate risk tolerance

$$t \equiv \rho\tau^\#.$$

The quadratic carry contract we propose gives them a payoff based on the fund's terminal value $V = \alpha^D \tilde{v}$ of the form

$$g(V) = \phi V + \frac{\beta}{2}(V - \bar{V})^2, \qquad \bar{V} = \alpha^D \mu, \quad \beta > 0. \tag{11}$$

Thus, they receive a linear share ϕ as in the main model, plus a convex carry measured relative to a benchmark (hurdle) $\bar{V}$ equal to the fund's *expected* value. The clients hold the residual claim $V - g(V)$. Two features of this contract help keep the analysis tractable and are economically natural. Using a quadratic carry helps us compute the FMs' certainty equivalent in closed form. Further, benchmarking this carry to the expected fund value $\bar{V}$ makes it mean-zero in payoff ($E[V - \bar{V}] = 0$) and, as shown below, monitoring-neutral.

1.2 The managers' certainty equivalent

The managers' certainty equivalent as a function of the fund's holding, α, is an exponential of a quadratic form in a Gaussian. If we write the carry in terms of the demeaned payoff $u \equiv \tilde{v} - \mu \sim N(0, \sigma^2)$, then, since $V - \bar{V} = \alpha^D u$ and $V = \alpha^D(\mu + u)$ (and using α for a general holding), then (11) can be expressed as

$$g(u) = \phi\alpha(\mu + u) + \frac{\beta}{2}\alpha^2 u^2 = \underbrace{\phi\alpha\mu}_{\text{constant}} + \underbrace{\phi\alpha}_{p}\, u + \underbrace{\tfrac{\beta}{2}\alpha^2}_{q}\, u^2. \tag{12}$$

Note that we treat the mean $\mu(m) = \mu$ as a constant since the monitoring choice is made prior to the realization of $\tilde{v}$. Thus

$$E\left[e^{-g(u)/t}\right] = e^{-\phi\alpha\mu/t}\, E\left[e^{\theta_1 u + \theta_2 u^2}\right], \quad \textit{where} \qquad \theta_1 = -\frac{p}{t} = -\frac{\phi\alpha}{t}, \quad \theta_2 = -\frac{q}{t} = -\frac{\beta\alpha^2}{2t}.$$

For $u \sim N(0, \sigma^2)$, we have

$$E\Big[e^{\theta_1 u + \theta_2 u^2}\Big] = \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{u^2}{2\sigma^2} + \theta_1 u + \theta_2 u^2\Big) du = \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-Au^2 + \theta_1 u\big) du,$$

with $A = \frac{1}{2\sigma^2} - \theta_2$. Completing the square and using $\int e^{-Au^2 + \theta_1 u} du = \sqrt{\pi/A}\, e^{\theta_1^2/4A}$ (which is valid iff $A > 0$), we have

$$E\left[e^{\theta_1 u+\theta_2 u^2}\right] = \frac{1}{\sqrt{2\pi\sigma^2}}\sqrt{\frac{\pi}{A}}\, e^{\theta_1^2/4A} = (1-2\theta_2\sigma^2)^{-1/2}\exp\Big(\frac{\frac{1}{2}\theta_1^2\sigma^2}{1-2\theta_2\sigma^2}\Big).$$

Now we define

$$D \equiv 1-2\theta_2\sigma^2 = 1+\frac{\beta\alpha^2\sigma^2}{t} > 1. \tag{13}$$

Then, with $\theta_1 = -\phi\alpha/t$ we have $\frac{1}{2}\theta_1^2\sigma^2 = \phi^2\alpha^2\sigma^2/2t^2$, so $E[e^{-g(u)/t}] = e^{-\phi\alpha\mu/t}D^{-1/2}\exp\big(\phi^2\alpha^2\sigma^2/2t^2 D\big)$.

Inserting this into the definition of the certainty equivalent net of monitoring costs, $CE_M = -t\ln\big(E(e^{-g(V)/t})\big) - c(m)$, we have

$$CE_M(\alpha) = \phi\alpha\mu + \frac{t}{2}\ln D(\alpha) - \frac{\phi^2\alpha^2\sigma^2}{2t\,D(\alpha)} - c(m), \qquad D(\alpha) = 1+\frac{\beta\alpha^2\sigma^2}{t}. \tag{14}$$

Note that at $\beta = 0$, $D = 1$ and (14) reduces to the ordinary CARA–normal value $\phi\alpha\mu - \frac{1}{2t}(\phi\alpha)^2\sigma^2$.

Monitoring Incentives. Because D does not contain μ, $\frac{\partial CE_M}{\partial\mu} = \phi\alpha$. The monitoring first-order condition $\phi\alpha\,\mu'(m) = c'(m)$ still sets the effective monitoring stake to $\phi\alpha$. Thus, imposing $\phi\alpha = \omega$ at the equilibrium gives $\phi = \phi^*$ and monitoring m^C, exactly as under the linear contract.

1.3 The trading equilibrium

Analogously to the main analysis, for global stability we require that the final chosen holding for the fund $\alpha^D = \alpha_G^D$ be a maximizer of $CE_M(\alpha^D) - CE_M(\alpha_G^D) - \phi(\alpha^D - \alpha_G^D)P^D(\alpha_G^D)$, with the price taken as given at its equilibrium value $P^D = P^D(\alpha_G^D)$. To satisfy the second criterion of globally stable allocation, α_G^D must also be a *global* maximizer of $CE_M(\alpha^D) - \phi\alpha^D P^D(\alpha_G^D)$ (see the Proof of Lemma **3**).

We first need $CE_M'(\alpha^D)$. From (13), $D'(\alpha^D) = \frac{2\beta\alpha^D\sigma^2}{t}$. Thus, we have

$$CE_M'(\alpha^D) = \phi\mu + \frac{\beta\alpha^D\sigma^2}{D} - \frac{\phi^2\alpha^D\sigma^2}{t\,D^2}. \tag{15}$$

With the market-clearing price $P^D(\alpha_G^D) = \mu - \dfrac{(1-\alpha_G^D)\sigma^2}{\rho(1-\lambda-\tau\#)}$, the first-order condition is $CE_M'(\alpha^D) = \phi P^D(\alpha_G^D)$. Dividing by $\phi\sigma^2$ and evaluating at the equilibrium $\alpha^D = \alpha_G^D$, we have

$$\frac{\phi^* \alpha_G^D}{t\, D^2} - \frac{\beta \alpha_G^D}{\phi^* D} = \frac{1-\alpha_G^D}{\rho(1-\lambda-\tau\#)}, \qquad D = 1 + \frac{\beta\, {\alpha_G^D}^2 \sigma^2}{t}. \tag{16}$$

To find the required convexity to achieve the targeted allocation, we must solve this equation for β with the targets $\phi = \phi^*$ and $\alpha_G^D = \omega + \alpha^C$. Note that at $\beta = 0$ ($D = 1$) this reduces to the equivalent expression for the main model. For $\beta > 0$, ((16)) is not analytically solvable for β, so we provide a numerical solution below.

Global stability is a binding constraint. The first global stability condition for the replication of the optimal allocation at the average value—a local critical point at $\alpha_G^D = \omega + \alpha^C$—can be satisfied well below τ^*. The second global stability condition, that α_G^D must also be a *global* maximizer of $CE_M(\alpha^D) - \phi\alpha^D P^D(\alpha_G^D)$, is more demanding, as it compares the equilibrium holding against far-away holdings, at each of which the managers re-optimize monitoring. Two forces destabilize $\alpha_G^D = \omega + \alpha^C$ as the global optimum. The *carry appetite*—the term $\beta\alpha\sigma^2/D$ in (15)—is a positive feedback: a larger holding raises D, which raises the value of the carry, encouraging a larger holding still. The *monitoring feedback* is a second: a larger holding induces more monitoring, which raises the asset's value $\mu(m)$, again encouraging accumulation. Both are versions of the endowment-effect force familiar from proprietary blockholding. Below a threshold $\tau^{\#}_{\min}$ the required degree of convexity implies that the *global* optimum jumps toward the competitive allocation, so the contract can no longer hold the fund at the optimum.

1.4 Numerical illustration

For our numerical illustration, we take $\lambda = 0.40$, $\omega = 0.20$, $\sigma^2 = 1$, $\rho = 1$, $\mu(m) = \mu_0 + m$ with $\mu_0 = 1$, and $c(m) = \frac{1}{2}\gamma m^2$ with $\gamma = 2$. Then $m^C = \omega/\gamma = 0.10$, $\mu(m^C) = 1.10$, $c(m^C) = 0.01$, and the benchmark allocation is $\tau^* = 0.1826$, $\phi^* = 0.3684$, $\alpha^C = 0.3429$, $\alpha^D = 0.5429$. Solving the trading equilibrium and locating the global-stability boundary gives a feasibility floor of $\tau^{\#}_{\min} \approx 0.136$, so the convex carry reproduces the client-optimal allocation while replacing up to about a quarter of the managers (from $\tau^* = 0.183$ down to 0.136). Table 2 below gives the required convexity (β) for different constrained levels of $\tau^{\#}$ between $\tau^*$and the feasibility floor $\tau^{\#}_{\min}$.

$\tau^{\#}$	$\beta(\tau^{\#})$
0.1826 ($= \tau^*$)	0.000
0.170	0.022
0.160	0.038
0.150	0.054
0.1358 ($= \tau^{\#}_{\min}$)	0.077

Table 2: Required carry across the feasible range.

2 Discussion: Concavity of $\Psi^D(\cdot)$.

The concavity of $\Psi^D(\alpha^D)$ as defined in (1), evaluated along the optimal monitoring path, requires that the curvature of the risk-bearing cost (the term $\frac{\phi^2(\alpha^D)^2\sigma^2}{2\rho\tau}$) dominates any convexity arising from the monitoring benefit $\phi\alpha^D\mu(m^D(\alpha^D))$. By the envelope theorem applied to the monitoring optimization, the second derivative of Ψ^D along the optimal path equals $\phi\mu'(m^D)m^{D\prime} - \frac{\phi^2\sigma^2}{\rho\tau}$, where $m^{D\prime} = \frac{\phi\mu'(m^D)}{c''(m^D)-\phi\alpha^D\mu''(m^D)}$. Sufficient conditions for strict concavity include $\mu''(\cdot) < 0$ and $c''(\cdot)$ sufficiently large relative to $\mu'(\cdot)^2$; the quadratic cost specification $c(m) = \frac{1}{2}\gamma m^2$ used in our numerical illustrations satisfies this for all standard parameterizations.