Is There Too Much Benchmarking in Asset Management?

Anil K Kashyap,* Natalia Kovrijnykh,† Jian Li,‡ and Anna Pavlova§

Abstract

We propose a tractable model of asset management in which benchmarking arises endogenously, and analyze its welfare consequences. Fund managers' portfolios are not contractible and they incur private costs in running them. Incentive contracts for fund managers create a pecuniary externality through their effect on asset prices. Benchmarking inflates asset prices and creates crowded trades. The crowding reduces the effectiveness of benchmarking in incentive contracts for others, which fund investors fail to account for. A social planner, recognizing the crowding, opts for contracts with less benchmarking and less incentive provision. The planner also delivers lower asset management costs.

JEL Codes: D82, D86, G11, G12, G23

Keywords: Incentive Contracts, Moral Hazard, Relative Performance, Pecuniary Externality, Asset Management, Benchmark, Index, Welfare

We have benefited from comments and suggestions by seminar and conference participants at Arizona State University, Chicago Booth, Columbia University, Federal Reserve Board of Governors, Higher School of Economics, INSEAD, London Business School, NYU Stern, University of Oxford, UK Financial Conduct Authority, Virtual Finance Theory Seminar, CERGE-EI Macro Workshop, JHU Carey Finance Conference, NBER Long-Term Asset Management meeting, NBES Conference, SAET in Ischia, SED in Minneapolis, SFS Cavalcade North America, WFA 2021, as well as comments of Adrian Buss, Hector Chade, Eduardo Davila, Alp Simsek, Dimitri Vayanos, Ming Yang, editors Pierre-Olivier Gourinchas and Sylvain Chassang, and three anonymous referees. This research has been supported by a grant from the Alfred P. Sloan Foundation to the Macro Financial Modeling project at the University of Chicago. The views expressed here are ours only and not necessarily those of the institutions with which we are affiliated.

*Booth School of Business, University of Chicago, National Bureau of Economic Research, and Centre for Economic Policy Research. Email: anil.kashyap@chicagobooth.edu.

†W. P. Carey School of Business, Arizona State University. Email: natalia.kovrijnykh@asu.edu.

‡Columbia Business School. Email: jl5964@columbia.edu.

§London Business School and Centre for Economic Policy Research. Email: apavlova@london.edu.

1 Introduction

Investors worldwide have delegated the investment of over $100 trillion to asset management firms. These firms then turn the decision over how to invest the money to portfolio managers, who have a principal-agent relationship with investors. Portfolio managers are invariably paid based on how their fund performs relative to a benchmark.[1] The presence of benchmarks in compensation contracts is important because benchmarks are a significant driver of global capital flows and have an effect on the real economy. For example, Calomiris et al. (2022) document that emerging market firms are able to cut their cost of funds by an astounding 1 percentage point by issuing bonds eligible for inclusion in important international benchmark indices. We provide a tractable model of asset management in which benchmarking arises endogenously. More importantly, we use our model to assess the welfare implications of benchmarking and explore its unintended consequences.

To study these issues, we embed an optimal-contracting problem in a general-equilibrium setting. We show that when the fund managers incur a private cost in managing portfolios, optimally designed contracts for the managers involve benchmarking. Because of this private cost, managers underinvest in the risky asset (stock market). Conditioning the managers' compensation on the performance of a benchmark portfolio partially protects them from risk and thus boosts their incentives to invest. In general equilibrium, the use of such incentive contracts creates a pecuniary externality through their effect on the risky asset's price. Benchmarking inflates the price of the risky asset and reduces its expected return. This in turn reduces the marginal benefit of using incentive contracts for others. We show that a constrained social planner, who internalizes this externality, would opt for less incentive provision and less benchmarking.

Here is how our model works. Some agents in the economy—direct investors—manage their own money and others—fund investors—delegate their invest-

[1] For example, Ma, Tang and Gómez (2019) report that around 80% of U.S. mutual funds explicitly base compensation on performance relative to a benchmark (usually a prospectus benchmark such as the S&P 500, Russell 2000, etc.).

ment choice to fund (or portfolio) managers. All agents are risk averse. Critically, the managers' portfolios are unobservable to fund investors and the cost of managing a portfolio is private. The managers are paid based on incentive contracts designed by the fund investors.[2] We focus on linear contracts, which include a fixed salary, a fee for absolute performance, and potentially a fee for performance relative to a benchmark.

We assume that the managers can potentially generate superior returns (or "alpha") relative to those of the direct investors through various sophisticated strategies. These include lending securities, conserving on transactions costs (e.g., from crossing trades in-house or by obtaining favorable quotes from brokers) or providing liquidity (i.e., serving as a counterparty to liquidity demanders and earning a premium on such trades). While these activities augment returns, they are associated with a private cost for a portfolio manager. We assume the costs are increasing in the size of the fund's risky portfolio. The simplest way to justify these assumptions is to appeal to the time costs involved in the activities and to interpret the rising costs as reflecting the additional time required for managing a larger fund/portfolio.

Fund investors design the manager's compensation contracts to incentivize the manager to take the risk associated with the sophisticated strategies. The presence of the private cost calls for a contract that rewards the manager based on fund performance and gives her a larger share of the return than if risk sharing were the only purpose of the contract. Because the stock market return is stochastic, rewarding performance exposes the manager to additional risk. This risk, if unmitigated, means that the manager will underinvest. Adding a benchmark to the contract partially protects the manager from this risk and therefore will be used by fund investors to improve the manager's incentives.

Our paper's main contribution is analyzing welfare consequences of benchmarking. When all fund investors use incentive contracts, they increase the

[2]We abstract from the asset management firm and assume that the firm acts in the interest of the fund investors, so that effectively the fund investors directly control the compensation arrangements for the portfolio managers. This is consistent with the fund trustees having a fiduciary obligation to their investors.

total demand for the risky asset. The increased demand boosts the price of the risky asset and lowers its expected return. In other words, benchmarking creates crowded trades.

Importantly, individual fund investors in our model take the stock price as given and do not internalize the effects of contracts they design on the equilibrium stock price. Crowded trades resulting from the contract-induced incentives are a pecuniary externality. Because of the agency frictions, markets are incomplete, so this pecuniary externality leads to an inefficiency. Specifically, the use of benchmarking contracts by a group of investors reduces the effectiveness of contracts designed by other investors through crowded trades. This happens because rewarding performance implies that the stock price enters the fund managers' incentive constraints. Each manager still has to incur the full private cost of managing assets but the benefits of doing so are reduced because of the crowded trades.

In light of this, it is natural to ask how would the incentive contract chosen by a social planner, who is subject to the same restrictions as individual investors but recognizes the effect of contracts on the stock price, differ from the privately optimal one? We show that individual investors underestimate the cost of incentive provision relative to the social planner, who internalizes the negative externality of incentive contracts. As a result, the planner opts for less incentive provision. Specifically, we show that both the performance sensitivity ("skin in the game") as well as the level of benchmarking are lower in the socially optimal contract than in the privately optimal one. This ameliorates the price pressure that portfolio managers exert and reduces the crowdedness of trades.

Our model informs the debate over whether the costs of asset management are excessive and whether returns delivered by the fund managers justify these costs. We use the model to compare the managers' costs and expected returns under privately and socially optimal contracts. We find that, from the socially optimal point of view, fund investors over-incentivize risk-taking so that managers invest too much at too high a cost.[3] In the equilibrium with privately

[3]While the cost is borne by the manager, it ultimately gets passed on to the fund

optimal contracts, the stock price is higher and consequently the expected per-share return is lower than under the socially optimal contract. Key to these implications is that, in contrast to fund investors, the planner internalizes the pecuniary externality arising from crowded trades.

The remainder of the paper is organized as follows. In the next section, we review the related literature. Section 3 presents the model. Section 4 analyzes the model and presents the main results. Section 5 concludes and outlines directions for future research. Omitted proofs, derivations, and other extensions are in the appendices.

2 Related Literature

Our work builds on the vast literature on optimal contracts with moral hazard. In a seminal contribution, Holmstrom (1979) argues that including a signal that is correlated with the output of the manager—in our case, the benchmark's performance—in a contract is beneficial to the principal. Importantly, in our paper the benefit of including the signal is endogenous through the general-equilibrium effect on the stock price. To our knowledge, ours is the first paper that endogenizes the effectiveness of including the extra signal in the contract. Holmstrom and Milgrom (1991) introduce a tractable contracting setting with moral hazard, with which our model shares many similarities, and show that increasing the agent's share in the project's output helps provide incentives. In the context of delegated asset management though, giving the agent a larger share of portfolio return encourages her to scale down the risk of the (unobservable) portfolio by reducing risky asset holdings. Stoughton (1993) and Admati and Pfleiderer (1997) show that the manager is able to completely "undo" her steeper incentives to collect information on asset payoffs by such scaling. We design a contract that overcomes this challenge and show that it involves benchmarking. Another notable difference from the aforementioned literature is that we embed optimal (linear) contracts in a general-equilibrium setting and study interactions between contracts and

investor, who needs to compensate the manager enough to ensure her participation.

equilibrium prices, and the implications of these interactions on welfare.

Our work is also related to the literature in asset pricing and corporate finance theory that explores the general-equilibrium implications of benchmarking. Brennan (1993) shows that benchmarking leads to lower expected returns on stocks included in the benchmark. In dynamic models, Cuoco and Kaniel (2011) and Basak and Pavlova (2013) show that benchmarking pushes up prices and lowers Sharpe ratios of stocks inside the benchmark. Basak and Pavlova also show that benchmarking leads to excess volatility and excess co-movement of returns on these stocks. Kashyap et al. (2021) focus on implications of benchmarking portfolio managers for firms' corporate decisions and demonstrate that firms in the benchmark have a higher valuation for investment projects or merger targets. These papers take the benchmarking contract of managers to be exogenous.

There are very few papers that study the asset pricing implications of relative performance evaluation in asset management with optimal contracts. Kapur and Timmermann (2005) analyze the effects of relative performance evaluation on the equity premium. In their paper, managers have exogenously superior information about assets compared to investors, and investors use contracts purely for risk-sharing purposes. In Buffa, Vayanos and Woolley (2014)[4] and Cvitanic and Xing (2018), benchmarking helps reduce diversion of cash flows by fund managers. Our rationale for benchmarking is to reward activities that generate superior returns. Sockin and Xiaolan (2020) study costly information acquisition by managers,[5] and, like us, highlight the pecu-

[4]In the published version, Buffa, Vayanos and Woolley (2022), constraints limiting deviations from benchmarks guard against the possibility that unskilled managers choose overly risky portfolios.

[5]See also Ozdenoren and Yuan (2017) who conduct a related analysis in the context of an industry equilibrium, in a classical moral-hazard setting with many principal-agent pairs. They show that benchmarking is privately optimal but it creates overinvestment and excessive risk-taking at the industry level. Albuquerque, Cabral and Guedes (2019) present a related model of industry equilibrium, enriched further with strategic interactions among firms in the industry, and show that benchmarking against peer performance induces agents to take correlated actions. Huang, Qiu and Yang (2020) analyze a model of delegated asset management with asymmetric information and endogenous contracts (but without relative performance) to study the effect of institutional investors on price informativeness. Unlike us, they limit their analysis to privately optimal contracts and do not study welfare

niary externality that emerges because of the effect of contracts on equilibrium prices. In contrast to us, they show that a constrained social planner opts for more incentive provision and more benchmarking.

Our paper also relates to the literature on pecuniary externalities in competitive equilibrium settings with incomplete markets.[6] Lorenzoni (2008) studies a model of credit booms in which a pecuniary externality arises from the combination of limited commitment and asset prices being determined in spot markets. Decentralized equilibria feature over-borrowing relative to the constrained optimum. Both our setting and mechanism are very different, but we share a similar prediction that asset prices in the decentralized equilibrium fall between those in the constrained and unconstrained optima. He and Kondor (2016) study a model in which individual firms' liquidity management decisions generate investment waves. These investment waves are constrained inefficient when future investment opportunities are noncontractible, and the social and private value of liquidity differs. In their model, overinvestment occurs during booms and underinvestment during recessions.

Gromb and Vayanos (2002) analyze a model in which competitive financially constrained arbitrageurs supply liquidity to the market, and fail to internalize the fact that their trading, in aggregate, affects prices. A social planner can achieve a Pareto improvement by either reducing or increasing the arbitrageurs' liquidity supply. Davila and Korinek (2018) highlight a distinction between "distributive externalities" that arise from incomplete insurance markets and "collateral externalities" that arise from price-dependent financial constraints. The externality in our paper falls into the second category, broadly defined, although in our case the inefficiency arises from the incentive problem rather than financial constraints. Di Tella (2019) studies optimal long-term contracts in a general-equilibrium model where financial interme-

---

implications. Donaldson and Piacentino (2018) propose a model in which a rationale for benchmarking in managers' contracts is to attract fund inflows. Dybvig, Farnsworth and Carpenter (2010) show that benchmarking emerges as optimal compensation in an environment where portfolio managers exert effort to improve the quality of a private signal about future prices.

[6] This literature goes back to Hart (1975), Greenwald and Stiglitz (1986), and Geanakoplos and Polemarchakis (1996).

diaries manage capital on behalf of households and can divert capital to sell for private gains. He shows that, due to a pecuniary externality, competitive equilibrium is not constrained efficient and the socially optimal allocation can be implemented with a tax on asset holdings.[7]

Biais, Heider and Hoerova (2021) analyze a model in which protection buyers trade derivatives with protection sellers and there is moral hazard on the side of protection sellers. In their model, although prices enter incentive constraints, a pecuniary externality does not lead to constrained inefficiency, as it does in our model, because investors can trade insurance against the risk of fire sales. We would have a similar result if we allowed for fully state-contingent contracts in our environment—see our discussion at the end of subsection 3.3. In Acemoglu and Simsek (2012), firms trade off providing insurance to workers and incentivizing them to exert effort. The authors show that, under certain conditions, equilibrium prices can tighten incentive constraints. They mainly focus on inefficient sharing of idiosyncratic risk. Instead, our focus is on the inefficient use of an additional signal—return of the benchmark portfolio—in the incentive contract.

There is some empirical evidence that benchmarking creates crowded trades. Lines (2016) observes that in times of high market volatility, portfolio tracking error rises. This leads portfolio managers to rebalance their portfolios towards benchmark stocks. He finds that this trading behavior leads to lower returns for the rebalanced portfolios.

3 Model

To illustrate our mechanism and main results in the simplest way, we set up a model with one risky asset. However, all the main results extend to the case with multiple risky assets—see Remark 5 at the end of Section 4.

[7]In a separate paper, Di Tella (2017) shows that there is another source of inefficiency if only short-term contracts are allowed.

3.1 Investment Opportunities and Agents

Except for portfolio managers and their clients, our environment is standard. There are two periods, $t = 0, 1$. Investment opportunities consist of a single risky asset (a stock or the stock market) and one risk-free bond. The stock is a claim to a cash flow $\tilde{D}$, realized at $t = 1$, where $\tilde{D} \sim N(\mu, \sigma^2)$. The risk-free bond pays an interest rate that is normalized to zero. There are $\bar{x} > 0$ shares of the risky asset and the bond is in infinite net supply. The stock price is denoted by p.

There is a continuum of agents of three types: direct investors, fund investors and fund managers. Direct investors manage their own portfolios. Fund investors can only buy the bond themselves and hire the managers to trade both the stock and bond on their behalf. Each manager works for one fund investor, and is restricted to invest her personal wealth in the bond. The fractions of direct investors and managers in the population are λ_D and λ_M, respectively, and the total population is normalized to one so that $\lambda_D + 2\lambda_M = 1$.

Each agent has a constant absolute risk aversion (CARA) utility function over final wealth (or compensation in the case of the manager) W, $U(W) = -e^{-\gamma W}$, where $\gamma > 0$ is the coefficient of absolute risk aversion. Direct investors and fund investors are endowed with x^D_{-1} and x^F_{-1} shares of the risky asset, respectively, where $\lambda_D x^D_{-1} + \lambda_M x^F_{-1} = \bar{x}$.[8]

We do not model an agent's choice to become a direct investor or a fund investor—the fractions of different investors in the population are exogenous. One could endogenize this choice, for example, by assuming heterogeneous costs of participating in the asset market. In Remark 4 at the end of Section 4 we describe the additional considerations that arise in this kind of extension, but we do not consider it here to maintain our focus on the central message of the paper.

[8]Without loss of generality, we assume that the managers are not endowed with the risky asset.

3.2 Value Added and Costs of Asset Management

For fund investors, delegating investment to a portfolio manager has costs and benefits. The benefits are that managers can potentially outperform direct investors. This advantage arises from having set up return-augmenting activities such as securities lending, providing liquidity by market making, or minimizing trade costs.[9]

In terms of the costs, delegation comes with an agency problem: the manager's portfolio choice is not contractible meaning that fund investors cannot write contracts that condition the manager's compensation directly on their portfolio choice. Non-contractibility can occur, for example, if the fund investors do not observe the manager's portfolio choice. This is a realistic assumption because even when managers are required to disclose their portfolios at particular points in time, their actual portfolios between the disclosure dates typically differ from their reported portfolios (Kacperczyk, Sialm and Zheng, 2008), and a fund investor cannot obtain detailed information on the manager's trades. Furthermore, the managers incur a private cost in managing a portfolio. For example, managers must monitor market conditions to successfully lend shares. In Online Appendix E.1, we also investigate an extension where the private cost is related to effort that cannot be observed. We elaborate on the interpretation of these benefits and costs in Online Appendix C.

We model the costs and benefits as follows. Throughout, we will work with per-share rather than per-dollar returns. The return for a direct investor's portfolio x is given by $x(\tilde{D} - p)$. The fund manager's return is

$$r_x = x(\Delta + \tilde{D} - p) + \varepsilon, \tag{1}$$

where $\Delta \geq 0$ is the (exogenous) expected abnormal return and $\varepsilon \sim N(0, \sigma_\varepsilon)$ is a noise term. We will refer to the excess return of $x\Delta + \varepsilon$ as "alpha." The

[9] Kashyap et al. (2022) includes an analysis of an alternative model in which the managers have stock-picking ability that comes from an informational advantage. That model is much more complicated, but we show that the mechanism is the same as in the model in this paper and the key results from this paper carry over.

manager incurs a private portfolio-management cost $x\psi$, where $\psi > 0$ is the exogenous cost per share.

There are several key ingredients that are crucial for our results. It is essential for our mechanism that the manager's portfolio is not contractible (or unobservable), and the manager incurs a private cost of managing it (meaning that this cost is borne by the manager and cannot be directly shared with the fund investor through the contract). This cost will lead to a misalignment of the fund investor's and management's preferences for the risky asset. If there were no costs (or if they could be passed on to the fund investor), there would be no incentive problem, and the results would be trivial.

The other key ingredient is the noise ε in the return-augmenting activities.[10] It exposes the managers to additional risk in their compensation.[11] While the fund investor can partly shield the manager from the dividend risk by benchmarking, this additional risk cannot be eliminated. As a result, contracts will fail to achieve first best.[12]

Unlike ψ and ε, the variable Δ is not essential for our results, and we include it only for realism. If the managers could not outperform the direct investors, there would be no justification to hire them. Nonetheless, if we ignore all the empirical evidence that suggests that asset manager can add value and set $\Delta = 0$, the incentive problem and risk-sharing problems would still be present, and all of our results would go through.

Finally, for simplicity, we assume that the fund's abnormal return Δ is exogenous, which means we are ignoring market participants who would be on the other side of the transaction. Presumably the other party would have an abnormal return of $-\Delta$ per share. In addition, one might argue that we are ignoring the effects of crowded trades on Δ. To formalize these considerations,

[10] One might wonder what happens if the noise is proportional to x (that is, the noise term is εx instead of ε). This is a special case of the extension that we analyze in Online Appendix E.1. The algebra is more involved in this case, but the main mechanism is the same.

[11] With one risky asset, ε also ensures that the investors cannot infer the exact portfolio choice of the manager from the observed return (as in Holmstrom and Milgrom, 1991). With multiple risky assets, it would not be possible to infer the portfolio even without the ε.

[12] We come back to the issue of why ε is needed in subsestion 4.3 following Lemma 2.

one needs to be more precise about the activity that generates Δ. Since we attempt to capture several of them, in the body of the paper we abstract from fully modeling any particular market. In Online Appendix E.2, we endogenize Δ and assume that it comes from securities lending. In this case, we show that when we account for the short sellers and endogenize Δ, all our major insights carry through.

3.3 Contracts

To provide incentives for the managers to invest in the risky asset and to generate alpha, the fund investors design compensation contracts. The managers receive compensation w from fund investors. We assume that this compensation has three parts: the first is a linear payout based on absolute performance of the manager's portfolio x, a second part that depends on the performance relative to a benchmark portfolio, and a third that is independent of performance.[13] The benchmark portfolio is one share of the risky asset. That is, the manager's compensation is given by

$$w = \hat{a} r_x + b(r_x - r_{\mathbf{b}}) + c = a r_x - b r_{\mathbf{b}} + c, \tag{2}$$

where r_x is the performance of the manager's portfolio defined in (1) and $r_{\mathbf{b}} = \tilde{D} - p$ is the performance of the benchmark portfolio.[14] The contract for a manager depends on three numbers $(\hat{a}, b, c)$—or, equivalently, (a, b, c). We refer to $\hat{a}$ as the sensitivity to absolute performance and b as the sensitivity to relative performance. Our main analysis and the intuitions that follow will be in terms of a rather than $\hat{a}$. We refer to the variable a as the manager's "skin in the game." The contract for a particular manager is optimally chosen by the fund investor who employs her. As we mentioned earlier, the manager is restricted to investing her personal wealth in the bond and so she cannot

[13]The third part captures features such as a fee linked to initial assets under management or a fixed salary or any fixed costs.

[14]Given that there is only one risky asset, we effectively normalize the benchmark portfolio to one share of the risky asset. In a general model with multiple risky assets, the benchmark portfolio is a vector.

"undo" her contract via trading in her personal account.[15]

We think of a manager's contract as a compensation contract between a portfolio manager and her investment-advisor firm (e.g., BlackRock, who we assume is acting in the interests of the fund investors). The structure of the contract in (2) is consistent with empirical evidence. For example, Ma, Tang and Gómez (2019) analyze mandatory disclosures by U.S. mutual funds and find that around 80% of the funds explicitly base managers' compensation on performance relative to a benchmark (usually the prospectus benchmark, e.g., S&P 500, Russell 2000, etc.). Managers also have a fixed salary component, but the fraction of fund managers whose entire compensation consists of only fixed salary is very small.[16]

The important feature of the contract driving our results is that the contract for the manager depends on the (per share) return, and hence the price of the risky asset. If asset prices vary across states (or time), then the compensation contract would necessarily depend on prices. Loosely speaking, if the fund investors were choosing for themselves, they would opt to buy less of the stock when its price is high. In delegating to the managers, the investors still want this consideration to be there. So, this feature of our contract is very realistic.

The restriction to linear contracts warrants some discussion. First, linear contracts make our model tractable and allow us to find optimal contracts in closed form. The closed-form solutions show the reader exactly where the various effects are coming from, and allow us to build intuition. However, our mechanism extends beyond the linear contracts considered here. The cen-

[15] In practice, portfolio managers have a fiduciary duty to their investors. This precludes them from taking actions that harm the investors, or engaging in any activity that creates a conflict of interest between the manager and the fund investors. Compliance departments at asset management firms attempt to deal with these problems by requiring pre-approval of many types of trades by the manager or banning them altogether, and restricting when trading can occur. A trade such as shorting a manager's benchmark would be blocked by these policies. (See U.S. Securities and Exchange Commission, 2004 for details.)

[16] The performance-based bonus exceeds the fixed salary for 68% of the funds in the Ma, Tang and Gómez (2019) sample, constituting more than 200% of fixed salary for 35% of funds. In contrast, Ibert et al. (2017) find surprisingly weak sensitivity of manager pay to performance for Swedish mutual funds.

tral results arise because the contracts raise the managers' demand, so that they will also drive up the equilibrium stock price. Individual investors do not account for this price effect but a social planner would recognize it. Consequently, a planner realizes that the price effect works against the incentive provision (as long as the manager's demand function is downward sloping) and will alter the contracts accordingly. This mechanism does not depend on contract linearity, and, intuitively, should be also present with other forms of contracts.

There is a subtle caveat, however, about the generality of the mechanism. The mechanism requires contracts not being fully state contingent/flexible. With fully state-contingent optimal contracts, the fund investors can effectively eliminate the dependence of the manager's incentive constraint on the price of the risky asset, which would yield to a constrained efficient outcome.[17] Nonetheless, our general mechanism would extend to environments with piecewise-linear contracts (e.g., "bonus" contracts of the form $w = \max\{ar_x - br_{\mathbf{b}}, 0\} + c$) or to cases in which contract parameters can differ across some but not all states.[18]

4 Analysis and Results

We now turn to the analysis of our model. We first present maximization problems of direct investors, fund managers, and fund investors. We then analyze privately and socially optimal contracts, and present the main results.

[17]As we discussed in the literature review, this result is akin to the finding in Biais, Heider and Hoerova (2021), who show that pecuniary externality does not lead to constrained inefficiency in their model because investors can trade insurance against the risk of fire sales. The result is different from that in Di Tella (2019), who finds that even with fully optimal contracts the decentralized equilibrium is constrained inefficient. The reason is that in his model the private benefit of diverting investment returns explicitly depends on the price. If we assumed that the private cost in our model includes the price of the risky asset, i.e., equal to $xp\psi$ instead of $x\psi$, then we would have the difference between privately and socially optimal contracts even with fully optimal contracts. For a broader analysis of issues arising in models with prices in incentive contracts see Kashyap et al. (2023*b*).

[18]The analysis of a discrete-state example with piecewise linear contracts, as well as the numerical analysis with bonus contracts (where we show numerically that our results hold) are available from the authors upon request.

4.1 Direct Investors' and Managers' Problems

At $t = 0$, each direct investor chooses the number of shares of stock, x, and risk-free bond holdings to maximize his expected utility $-Ee^{-\gamma W}$. Since his return on the portfolio is $x(\tilde{D}-p)$, the resulting time-1 wealth is $W = x^D_{-1}p + x(\tilde{D}-p)$. It is well known that with the CARA utility function and normally distributed returns, a direct investor's maximization problem is equivalent to the following mean-variance optimization: $\max_x \; x(\mu - p) - \gamma x^2\sigma^2/2$.

Next, consider the problem of a portfolio manager. Each manager chooses the number of shares of stock x and the risk-free bond holdings to maximize $-E\exp\{-\gamma[ar_x - br_{\mathbf{b}} + c - \psi x]\}$, where the quantity inside the square brackets is her compensation net of the private cost. This maximization problem is equivalent to the following mean-variance optimization:

$$\max_x \; ax(\Delta - \psi/a + \mu - p) - b(\mu - p) + c - \frac{\gamma}{2}\left[(ax-b)^2\sigma^2 + a^2\sigma_\varepsilon^2\right].$$

Note that the manager receives a fraction a of the per-share abnormal return on the assets, Δ, but pays the entire cost ψ per share. (We later show that $a < 1$.)

Both the direct investors and managers take the stock price as given. Lemma 1 reports the optimal portfolio choices of the direct investors and managers arising from their optimizations, and the market-clearing asset price (for a given contract) arising from the market-clearing condition $\lambda_M x^M + \lambda_D x^D = \bar{x}$.[19]

Lemma 1 (**Portfolio Choices and Market-Clearing Price**). *For a given a contract* (a, b, c), *(i) the direct investors' and managers' optimal portfolio*

[19] We define the equilibrium at the end of subsection 4.2 after we introduce the fund investor's problem.

choices are as follows:

$$x^D = \frac{\mu - p}{\gamma\sigma^2}, \tag{3}$$

$$x^M = \frac{\Delta - \psi/a + \mu - p}{a\gamma\sigma^2} + \frac{b}{a} = \frac{x^D}{a\gamma} + \frac{\Delta - \psi/a}{a\gamma\sigma^2} + \frac{b}{a}; \tag{4}$$

(ii) the market-clearing price of the risky asset is

$$p = \mu - \gamma\sigma^2\Lambda\left(\bar{x} - \lambda_M\frac{b}{a}\right) + \Lambda\frac{\lambda_M}{a}\left(\Delta - \frac{\psi}{a}\right), \tag{5}$$

where $\Lambda \equiv [\lambda_M/a + \lambda_D]^{-1}$ *modifies the market's effective risk aversion.*

A direct investor's portfolio is the standard mean-variance portfolio, scaled by his risk aversion γ. A manager's portfolio choice differs from that of a direct investor in three respects. First the manager holds the same scaled mean-variance portfolio, but because she only receives a of any performance that she generates, she adjusts her holdings by $1/a$. Second, because managers have access to return-augmenting strategies, they perceive the mean-variance trade-off differently from the direct investors and tilt their mean-variance portfolios to try to produce alpha. Consistent with this result, Johnson and Weitzner (2019) report that fund managers' portfolios in their sample overweight assets with high securities-lending fees. Finally, because the manager's compensation is exposed to fluctuations in the benchmark, she holds a hedging portfolio that is (in this case perfectly) correlated with the benchmark, i.e., the benchmark itself.[20] The split between the mean-variance portfolio and the benchmark is governed by the strength of the relative-performance incentives, captured by b. The higher the b, the closer the manager's portfolio to the benchmark.

Because contracts change the managers' demand functions, the equilibrium stock price will depend on these contracts. Benchmarking pushes up the stock

[20]This implication is very general, and we share it with other models that analyzed benchmarking, both in two-period and multi-period economies and for other investor preferences specifications. This result first appeared in Brennan (1993) in a two-period model. Cuoco and Kaniel (2011) and Basak and Pavlova (2013), among others, obtain it in dynamic models with different preferences.

price, thus lowering the stock's expected return. Unlike the social planner, individual fund investors take the stock price as given and do not account for this pecuniary externality. We turn to the fund investors' problem next.

4.2 Fund Investors' Problem

Each fund investor chooses a contract (a, b, c) and portfolio $x = x^M$ to maximize his expected utility subject to the manager's participation and incentive constraints. The latter is the manager's first-order condition (4), capturing the fact that the portfolio x is the manager's private choice.[21]

To write the fund investor's problem formally, it is convenient to express payoffs in terms of the following variables:

$$y = ax - b, \; z = x - y.$$

These are the effective allocations of asset holdings to the manager and fund investor, respectively. Then the fund investor's and manager's utilities (in the mean-variance form) can be written as follows:

$$U^F\left(a, \frac{b}{a}, c, y, p\right) = x(1-a)\Delta + z(\mu - p) - \frac{\gamma}{2}\left[z^2\sigma^2 + (1-a)^2\sigma_\varepsilon^2\right] - c + x_{-1}^F p,$$
$$U^M\left(a, \frac{b}{a}, c, y, p\right) = x(a\Delta - \psi) + y(\mu - p) - \frac{\gamma}{2}\left[y^2\sigma^2 + a^2\sigma_\varepsilon^2\right] + c,$$

where x and z are given by

$$\begin{aligned} x &= \frac{y}{a} + \frac{b}{a}, \\ z &= \left(\frac{1}{a} - 1\right) y + \frac{b}{a} = \left(\frac{1}{a} - 1\right) \frac{\Delta - \psi/a + \mu - p}{\gamma\sigma^2} + \frac{b}{a}. \end{aligned} \tag{6}$$

[21] We show in the proof of Lemma 1 that the manager's second-order condition is satisfied, and thus the first-order approach is valid.

Then the fund investor's problem can then be written as follows:[22]

$$\max_{a,b/a,c} \quad U^F$$
$$\text{s.t.} \quad U^M \geq u_0, \tag{7}$$
$$y = \frac{\Delta - \psi/a + \mu - p}{\gamma\sigma^2}. \tag{8}$$

Constraint (7) is the manager's participation constraint, where u_0 is (the mean-variance equivalent of) the value of manager's outside option.[23] Equation (8) is the manager's (modified) incentive constraint.

An equilibrium with privately optimal contracts consists of the contract, risky asset holdings by direct investors and fund managers, and the stock price such that the agents solve their corresponding problems and the stock market clears. Appendix A contains the formal definition. We characterize this equilibrium in the next subsection.

4.3 Privately Optimal Contracts

As a point of reference, consider the first best where the manager's portfolio choice is observable and contractible. The first best involves efficient risk sharing between the (equally risk-averse) fund investor and manager, and the contract that implements it is $a = 1/2$ and $b = 0$.[24]

However, if under efficient risk sharing the manager chose the portfolio privately, she would underinvest in the risky asset. A higher a reduces the manager's effective cost ψ/a, which increases her demand for the risky asset. However, a higher a also exposes the manager to more risk, which makes

[22]The formulation of the fund investor's problem in terms of the exponential utilities (rather than in the mean-variance form) can be found in Online Appendix B.

[23]We do not model explicitly what this outside option is, as it does not matter for our main results. It can be exogenous, or it can be endogenized. Notice also that because of the contract's constant component c, in the mean-variance formulation utility becomes transferable, and the fund investor effectively maximizes the total utility of the fund investor and the manager subject to the manager's incentive constraint. The manager's participation constraint is then trivially satisfied by adjusting the constant c.

[24]See Lemma 5 in Appendix A for the formal analysis.

her scale down x^M, as can be seen in the denominator(s) of (4). Thus the use of performance pay creates a tension between incentive provision and risk sharing. The use of benchmarking, alleviates this tension by mitigating the adverse effect of a. Benchmarking shields the manager from risk by reducing variance in her compensation for a given portfolio choice.[25] As a result, (for the same a) the manager invests more. In what follows, we will consider how the fund investor will optimally choose the levels of a and b.

Notice that the fund investor fully internalizes the manager's cost of managing the fund.[26] But since the manager bears the cost privately and only receives fraction a of the return, for her the effective cost is higher, which is why ψ/a appears in (8). The difference between the full social cost and the cost perceived by the manager, ψ/a and ψ will be play an important role in the tradeoff between risk sharing and incentive provision.

First, consider the fund investor's optimal choice of relative performance in the contract, b. Notice that b enters into the fund investor's and manager's problems only though b/a. The first-order condition with respect to b/a is given by[27]

$$\frac{\partial(U^F + U^M)}{\partial(b/a)} = \Delta - \psi + \mu - p - \gamma\sigma^2 z = 0. \tag{9}$$

This condition captures the fact that an increase in b/a makes the manager invest more in the risky stock. Therefore, the optimal level of b will be the one that balances a marginal increase in the mean of the total expected surplus, $\Delta - \psi + \mu - p$, with the marginal increase in the variance, $\sigma^2 z$.

[25] By reducing the manager's risk exposure, benchmarking makes it cheaper for the fund investor to implement any particular portfolio choice.

[26] Formally, this can be seen by taking the first-order condition with respect to c, which implies that the Lagrange multiplier on the participation constraint equals one.

[27] We show in Lemma 7 in Appendix A that the second-order conditions hold in both privately and socially optimal cases.

Substituting out z using (6), equation (9) can be rewritten as[28]

$$\gamma\sigma^2 b = (2a-1)\left(\Delta - \psi + \mu - p\right) + (1-a)\left(\frac{1}{a} - 1\right)\psi. \tag{10}$$

The two terms on the right-hand side of equation (10) capture two considerations that fund investors have in mind when designing the benchmark. Note two extreme cases: $a = 1/2$ when perfect risk sharing is achieved, and $a = 1$ when the private and social costs are aligned. As we will show later, in the optimal contract $a \in (1/2, 1)$, so both terms on the right-hand side of (10) are positive. The first term, $(2a-1)(\Delta - \psi + \mu - p)$, arises because the fund investor recognizes that benchmarking increases the total expected surplus net of cost. Since $a > 1/2$, the manager is exposed to more risk than is efficient, so the fund investor uses benchmarking to make her invest more. The second term, $(1-a)(1/a-1)\psi$, reflects the incentive-provision role of b. By protecting the manager from risk, benchmarking provides her with incentives to invest more.

Notice that (10) depends on the equilibrium price p. When choosing b, the fund investor takes p as given. In equilibrium, however, p depends on the contract as given by equation (5). Then, to find the equilibrium value of b (the fixed point), we need to substitute (5) in (10) and solve it for b. This leads us to equation (13) in Lemma 2 below, which presents b only in terms of model parameters and a, which we will now solve for.

The first-order condition with respect to a is given by

$$\begin{aligned} 0 &= \frac{\partial(U^F + U^M)}{\partial a} + \frac{\partial U^F}{\partial y}\frac{\partial y}{\partial a} \\ &= -(2a-1)\gamma\sigma_\varepsilon^2 - \underbrace{\left(\Delta - \psi + \mu - p - \gamma\sigma^2 z\right)}_{=0 \text{ by FOC wrt } b/a,\ (9)}\frac{y}{a^2} + \frac{1-a}{a}\underbrace{\left(\Delta + \mu - p - \gamma\sigma^2 z\right)}_{=\psi \text{ by FOC wrt } b/a,\ (9)}\frac{\partial y}{\partial a} \\ &= -(2a-1)\gamma\sigma_\varepsilon^2 + (1-a)\frac{\psi^2}{\gamma\sigma^2 a^3}, \end{aligned} \tag{11}$$

where the last equality uses the first-order condition with respect to b/a, (9),

[28] See the proof of Lemma 2 in Appendix A for derivations.

and $\partial y/\partial a = \psi/(\gamma\sigma^2 a^2)$. First, notice the appearance of $\partial y/\partial a$. It captures how a marginal increase in a affects the manager's incentive to invest in the risky asset. This is the way that the contract creates incentives. Second, several terms drop out because b/a is chosen optimally, leaving only a term that is proportional to σ_ε^2. The cancellation comes because the optimal level of benchmarking already optimally shares the dividend risk, so all that remains to be shared is the extra risk from return-augmenting activities.

Notice that unlike in (10), the incentive-provision term and the risk-sharing term have different signs. This means that there is a tradeoff between incentive provision and risk sharing. A higher a is beneficial as it provides incentives for alpha-production, but is also costly because it exposes the manager to too much risk.

The following lemma summarizes the closed-form expressions for the equilibrium contract, as well as the expressions for the equilibrium price and stock holdings:

Lemma 2. *In the equilibrium with privately optimal contracts,*
(i) a^* and b^* solve

$$0 = (1-a^*)\frac{\psi^2}{\gamma\sigma^2 a^{*3}} - (2a^*-1)\gamma\sigma_\varepsilon^2, \tag{12}$$

$$b^* = (2a^*-1)\left[\bar{x} + \frac{\lambda_D}{\gamma\sigma^2}(\Delta-\psi)\right] + (1-a^*)\left[\frac{1}{a^*} - \left(\frac{\lambda_M}{a^*} + \lambda_D\right)\right]\frac{\psi}{\gamma\sigma^2}, \tag{13}$$

(ii) the risky asset's price is

$$p^* = \mu - \gamma\sigma^2\bar{x} + \lambda_M\left(2\Delta - \psi - \frac{\psi}{a^*}\right), \tag{14}$$

and each fund's risky asset holdings are

$$x^{M*} = 2\bar{x} + \frac{\lambda_D}{\gamma\sigma^2}\left(2\Delta - \psi - \frac{\psi}{a^*}\right). \tag{15}$$

Notice that there is a recursive structure to these conditions. The expression in (12) does not depend on b^* and is a function of only a^* and the model parameters.[29] Given a^*, (13), (14), and (15) deliver the expressions for b^*, p^*, and x^{M*}, respectively.

Let us briefly comment on the expression for the equilibrium price given by (14). Absent fund managers, the equilibrium price would be $p = \mu - \gamma\sigma^2\bar{x}$. The last term in (14) means that the price reflects the managers' extra demand associated with their return-augmenting activities. Notice that the term in parentheses is a sum of $\Delta - \psi$ and $\Delta - \psi/a$, which are the (marginal) extra expected returns net of costs as perceived by the fund investors and by the managers, respectively. Similarly, the equilibrium asset holdings of managers in (15) are higher when the opportunities for alpha-production are better. Notice that managers hold exactly $2\bar{x}$ when $\lambda_D = 0$. We will discuss this special case further in subsection 4.4.

Next, we turn to the characterizations of a^* and b^*, (12) and (13). We prove below that the equilibrium level a^* is strictly between 1/2 (perfect risk sharing) and 1 (private and social costs coincide). Also note that as σ_ε^2 goes to zero, a^* approaches 1, and the allocation approaches the first-best one (see Lemma 5 in Appendix A.) Indeed, it is crucial for our results that the fund investor does not "sell the project" to the manager, i.e., $a^* < 1$. As an alternative to the assumption of $\sigma_\varepsilon^2 > 0$, there are other modeling choices that would ensure that $a^* < 1$, for example, a lower-bound on c, the constant part of the contract.

As long as $\bar{x} + \lambda_D(\Delta - \psi)/(\gamma\sigma^2) > 0$, all the terms on the right-hand side of (13) are positive. This condition is satisfied either if $\Delta - \psi > 0$ (the expected abnormal return exceeds the cost of managing the portfolio), or if the net supply of the stock $\bar{x}$ is large enough. This brings us to our first main result.

Proposition 1 (**Benchmarking is Optimal**). *Consider the equilibrium with privately optimal contracts.*
(i) The equilibrium value of the "skin in the game" satisfies $a^* \in (1/2, 1)$.
(ii) Suppose that $\bar{x} + \lambda_D(\Delta - \psi)/(\gamma\sigma^2) > 0$. *Then there is benchmarking, that*

[29] Equation (12) has two roots, one positive and one negative. The negative root can be ruled out by the manager's second-order condition, see the proofs of Lemma 1 and Proposition 1 (i).

is, $b^* > 0$*.*

Part (ii) of Proposition 1 is essentially a version of Holmstrom's (1979) famous sufficient-statistic result—the use of an additional signal (in this case, the benchmark return) helps the contract designer provide incentives to the manager in a more effective way. While Holmstrom's result suggests that b^* is different from zero in general, provided $\bar{x} + \lambda_D(\Delta - \psi)/(\gamma\sigma^2) > 0$ we can say b^* is strictly positive, which is the relevant case given this application.

This proposition helps us understand why benchmarking in the asset management industry is so pervasive. Benchmarking is useful to fund investors because it incentivizes the manager to engage more in risky return-augmenting activities by partially protecting her from risk. In the language of the asset management industry, benchmarked managers are being protected from "beta" (i.e., the fluctuations in the return of the benchmark/market portfolio) while being rewarded for "alpha."

We wrap up this subsection by stating some comparative-statics results:

Lemma 3 (**Comparative Statics**)**.** *Consider the equilibrium with privately optimal contracts.*
(i) If the cost of managing the fund portfolio, ψ*, is higher, then* a^* *is higher and* p^* *and* x^{M*} *are lower.*
(ii) If the expected excess return, Δ*, is higher, then* b^**,* p^* *and* x^{M*} *are higher.*
(iii) If the extra risk associated with producing excess returns, σ_ε^2 *is higher, then* a^**,* p^* *and* x^{M*} *are lower.*[30]

These results are intuitive. The higher the ψ, the more costly it is to incentivize the manager. The fund investor will react to an increase in the cost by giving the manager a larger share of the return. With a higher cost (and despite a higher a, since ψ/a^* is still increasing in ψ), the manager will invest less in the risky asset, leading to a lower stock price. On the other hand, the higher the extra risk associated with producing excess returns, σ_ε^2, the more important is the risk sharing. The fund investor will choose a lower a (closer to 1/2), giving the manager lower incentives to invest in the risky asset,

[30]Notice that the effects of ψ and σ_ε^2 on b^* are ambiguous.

again leading to lower x^{M*} and p^*. Finally, when Δ increases, the abnormal return is higher. As a result, the fund investors use more benchmarking to shield the managers from risk, so that the managers invest more in the risky asset.

4.4 Socially Optimal Contracts

Fund investors design contracts to influence the manager's demand for the risky asset. Through the collective demand of the managers, contracts influence the equilibrium stock price, as given by (5). The price then affects the marginal cost/marginal benefit tradeoff of contracts for all fund investors. Since fund investors take the stock price as given, they do not internalize how their choices of contracts (once aggregated) change the effectiveness of other fund investors' contracts. In other words, fund investors impose an externality on each other through their use of contracts. In this subsection, we ask what contract a planner, who is subject to the same restrictions as fund investors, would choose to internalize this externality.

We define the problem of a constrained social planner as follows. The planner maximizes the weighted average of fund investors' and direct investors' utilities subject to the participation and incentive constraints of the managers, as well as the constraint that direct investors choose their portfolios themselves.[31] As before, this problem can be equivalently rewritten in terms of mean-variance preferences.[32] Define $U^D = x^D_{-1}p + x^D(\mu - p) - \gamma x^{D2}\sigma^2/2$. Then the social planner's problem is

$$\max_{a,b/a,c} \ \omega_F U^F + \omega_D U^D$$

[31]Equivalently, instead of imposing the manager's participation constraint, her utility can be included in the planner's objective function with a Pareto weight ω_M. For the transfer c to be finite, we must have $\omega_M = \omega_F$. This is analogous to noticing that the Lagrange multiplier on the participation constraint, which effectively acts as the Pareto weight on the manager, equals ω_F.

[32]We provide the original formulation in terms of exponential utilities in Online Appendix B.

subject to (3), (5), (7), and (8).

The social planner's first-order condition with respect to b/a is

$$0 = \underbrace{\left[\omega_F \left(x_{-1}^F - x^M\right) + \omega_D \left(x_{-1}^D - x^D\right)\right] \frac{\partial p}{\partial (b/a)}}_{\text{distributive pecuniary externality}} + \omega_F \left[\underbrace{\frac{\partial (U^F + U^M)}{\partial (b/a)}}_{\text{private FOC}} + \underbrace{\frac{\partial U^F}{\partial y} \frac{\partial y}{\partial p} \frac{\partial p}{\partial (b/a)}}_{\substack{\text{contracting} \\ \text{pecuniary externality}}} \right]. \tag{16}$$

The terms in the first line of (16) capture what Davila and Korinek (2018) call "distributive effects" or "distributive pecuniary externality." Depending on the initial endowments and the Pareto weights, the social planner has incentives to use benchmarking to move the price so as to benefit one or the other party based on this distributive motive. We discuss the distributive effects in Remark 1 at the end of the next section. Our focus is on the "contracting pecuniary externality" that acts through the price entering the manager's incentive constraint. To isolate the planner's motive to correct the contracting externality from the distributive motive, we want to neutralize the latter. To do this, we set the Pareto weights equal to the population weights, $\omega_F = \lambda_M$ and $\omega_D = \lambda_D$.[33] Then by market clearing, $\omega_F \left(x_{-1}^F - x^M\right) + \omega_D \left(x_{-1}^D - x^D\right) = 0$, so the term in the first line of (16) is zero. (See Davila and Korinek, 2018 for further discussion.)

Rewriting the term in the second line of (16) yields

$$0 = \underbrace{\left(\Delta - \psi + \mu - p - \gamma\sigma^2 z\right) \underbrace{\frac{\partial y}{\partial (b/a)}}_{=1}}_{\text{private FOC}} + \underbrace{\frac{1-a}{a} \left(\Delta + \mu - p - \gamma\sigma^2 z\right) \frac{\partial y}{\partial p} \frac{\partial p}{\partial (b/a)}}_{\text{contracting pecuniary externality}}. \tag{17}$$

[33] Choosing Pareto weights to cancel out the distributive effects is equivalent to allowing the social planner to use transfers for any Pareto weights. The planner would then use transfers to equate the marginal utilities (weighted by Pareto weights) of different agents.

Compare (17) with the first-order condition with respect to b/a in the private case, (9). The first term in (17) is exactly (9). The second term in (17) captures the contracting pecuniary externality that the planner is trying to correct and that the private agents ignore.

Consider the term $[\partial y/\partial p][\partial p/\partial(b/a)]$. The term $\partial y/\partial p = -1/(\gamma\sigma^2)$ captures the fact that the manager's demand function is downward sloping. The term $\partial p/\partial(b/a) = \gamma\sigma^2\Lambda\lambda_M$ reflects the fact that the higher the value b/a collectively used by all fund investors, the more crowded the trades, and the higher the stock price. The product of the two, $[\partial y/\partial p][\partial p/\partial(b/a)] = -\Lambda\lambda_M = -\lambda_M/(\lambda_M/a+\lambda_D)$ captures the fact that the general equilibrium effect of contracts on the risky asset's price reduces the effectiveness of b/a in incentivizing the manager to hold more of the risky asset. Hence (17) becomes

$$\left(\Delta + \mu - p - \gamma\sigma^2 z\right)\left[1 - \frac{(1-a)\lambda_M/a}{\lambda_M/a + \lambda_D}\right] - \psi = 0, \tag{18}$$

or

$$\Delta - \underbrace{\frac{\lambda_M/a + \lambda_D}{\lambda_M + \lambda_D}\psi}_{\substack{\text{cost from the}\\ \text{planner's perspective}}} + \mu - p - \gamma\sigma^2 z = 0. \tag{19}$$

Similar to the fund investors, the planner trades off the benefits and costs of inducing the agent to invest in the risky asset. Fund investors think of the benefit as the usual mean-variance consideration given by $(\Delta + \mu - p - \gamma\sigma^2 z)$, and the cost as ψ. For the planner, the benefit is smaller than for the fund investors, because she realizes that benchmarking inflates the risky asset's price and thus reduces its expected return. Put differently, due to this crowded-trades effect, the cost is higher for the same units of benefit: the cost is $(\lambda_M/a + \lambda_D)/(\lambda_M + \lambda_D)\psi$ in (19) vs. ψ in (9). (This difference in the perceived costs will show up in our further comparisons between the socially and privately optimal contracts.) So, from the planner's point of view, incentive provision is less beneficial/more expensive, which, as we will see, will make her do less of it.

Substituting for z, we obtain the planner's counterpart to equation (10):

$$\gamma\sigma^2 b = (2a-1)\left[\Delta - \frac{\lambda_M/a+\lambda_D}{\lambda_M+\lambda_D}\psi + \mu - p\right] + (1-a)\left[\frac{1}{a} - \frac{\lambda_M/a+\lambda_D}{\lambda_M+\lambda_D}\right]\psi. \tag{20}$$

Compared to (10), the cost ψ is again replaced with $(\lambda_M/a+\lambda_D)/(\lambda_M+\lambda_D)\psi$. Finally, substituting in the equilibrium price, (5), yields the fixed-point value of b that depends only on the model parameters, as presented in equation (24) in Lemma 4 below.

Next, consider the planner's first-order condition with respect to a:

$$\begin{aligned} 0 = {} & \frac{\partial(U^F+U^M)}{\partial a} + \frac{\partial U^F}{\partial y}\left[\frac{\partial y}{\partial a} + \underbrace{\frac{\partial y}{\partial p}\frac{\partial p}{\partial a}}_{\substack{\text{contracting}\\ \text{externality}}}\right] \\ = {} & -(2a-1)\gamma\sigma_\varepsilon^2 - (\Delta - \psi + \mu - p - \gamma\Sigma z)\frac{y}{a^2} \\ & + \frac{1-a}{a}(\Delta + \mu - p - \gamma\Sigma z)\left[\frac{\partial y}{\partial a} + \frac{\partial y}{\partial p}\frac{\partial p}{\partial a}\right], \\ = {} & -(2a-1)\gamma\sigma_\varepsilon^2 + \frac{1-a}{a}\frac{\lambda_M/a+\lambda_D}{\lambda_M+\lambda_D}\psi\left[\frac{\partial y}{\partial a} + \frac{\partial y}{\partial p}\frac{\partial p}{\partial a} + \frac{y}{a^2}\frac{\partial y}{\partial p}\frac{\partial p}{\partial(b\theta/a)}\right], \end{aligned} \tag{21}$$

where the last equality follows from (17). After some algebra (see the proof of Lemma 4 in Appendix A) this condition can be written as follows:

$$-(2a-1)\gamma\sigma_\varepsilon^2 + (1-a)\frac{\psi^2}{\gamma\sigma^2 a^3}\frac{\lambda_D}{\lambda_M+\lambda_D} = 0. \tag{22}$$

Compare this equation to its analog in the case with privately optimal contracts, (11). Notice that the benefit of incentive provision captured by the first term in (22) is smaller than the corresponding term in (11). As a result, the planner will choose a lower a than fund investors will. We will formalize this result later in Proposition 2.

The following lemma presents the resulting expressions for the equilibrium contract, price, and risky-asset holdings in closed form.[34]

Lemma 4. *In the equilibrium with socially optimal contracts,*
(i) a^{**} and b^{**} solve[35]

$$0 = (1 - a^{**}) \frac{\psi^2}{\gamma \sigma^2 a^{**3}} \frac{\lambda_D}{\lambda_M + \lambda_D} - (2a^{**} - 1)\gamma \sigma_\varepsilon^2, \tag{23}$$

$$b^{**} = (2a^{**} - 1)\left[\bar{x} + \frac{\lambda_D}{\gamma\sigma^2}(\Delta - \psi)\right] + (1 - a^{**})\left[\frac{1}{a^{**}} - \frac{\lambda_M / a^{**} + \lambda_D}{\lambda_M + \lambda_D}\right]\frac{\psi}{\gamma\sigma^2}, \tag{24}$$

(ii) the risky asset's price is

$$p^{**} = \mu - \gamma\sigma^2\bar{x} + \lambda_M\left(2\Delta - \frac{\lambda_M/a^{**} + \lambda_D}{\lambda_M + \lambda_D}\psi - \frac{\psi}{a^{**}}\right), \tag{25}$$

and each fund's risky asset holdings are

$$x^{M**} = 2\bar{x} + \frac{\lambda_D}{\gamma\sigma^2}\left(2\Delta - \frac{\lambda_M/a^{**} + \lambda_D}{\lambda_M + \lambda_D}\psi - \frac{\psi}{a^{**}}\right). \tag{26}$$

Equations (23)–(26) are the analogs of (12)–(15) and have the same recursive structure. As expected, the two sets of equations coincide when $\lambda_M = 0$, and hence there is no externality. But so long as there are managers, the socially and privately optimal contracts are different. Proposition 2 below reveals how exactly they compare to each other.

We are now ready to present the central result of the paper.

Proposition 2 (Socially vs. Privately Optimal Contracts). *Compared to the privately optimal contract, the socially optimal contract involves*
(i) less "skin in the game," that is, $a^{**} < a^*$;
(ii) less benchmarking, that is, $b^{**} < b^*$, if $\bar{x} + \lambda_D(\Delta - \psi)/(\gamma\sigma^2) > 0$.[36]

---

[34] See Appendix A for the formal definition of the equilibrium with socially optimal contracts.

[35] From (23), $1/2 \le a^{**} < 1$, where the first inequality is strict so long as $\lambda_D > 0$.

[36] We also show in the proof of Proposition 2 that $b^{**}/a^{**} < b^*/a^*$.

As we have seen in our analysis, the use of contracts inflates the risky asset's price and thus *reduces the marginal benefit of incentive provision* for everyone else. The social planner internalizes this effect, and opts for less incentive provision than fund investors.

As a special case that helps make the point very clearly, suppose there are no direct investors, $\lambda_D = 0$. In this case, each fund will hold exactly $2\bar{x}$ shares and the total alpha in the economy is fixed, equal to $2\bar{x}\Delta$. The planner understands that incentive provision is unnecessary in this case, so there is no tradeoff between incentive provision and risk sharing. Indeed, by substituting $\lambda_D = 0$ into (23)−(24), it immediately follows that the socially optimal contract is $a = 1/2$ and $b = 0$, which coincides with the first-best one (see Lemma 5 in Appendix A). In contrast, the fund investors do not appreciate the fact that, in equilibrium, their contracts will not help them generate higher returns, and use contracts with $a > 1/2$ and $b > 0$, as can be seen from (12)−(13).

To further emphasize that benchmarking is crucial for the comparison between privately and socially optimal contracts, consider a case where benchmarking is exogenously set to zero, $b = 0$. In this case, all incentive provision and risk sharing has to be done through a. As we discussed earlier, an increase in a has two opposing effects on the managers' demands and hence the risky asset's price. It can be shown that with $b = 0$ the comparison between a^* and a^{**} is ambiguous. Intuitively, both the marginal benefit of a (incentive provision) as well as its marginal cost (exposing the manager to more risk) are lower for the social planner who internalizes the effect of a on the price. Depending on which reduction is bigger, the planner might choose a higher or a lower a than the fund investors do. Thus, only because of benchmarking ($b \neq 0$) can we be sure of the direction of the externality and are able to say that privately optimal contracts deliver excessive incentive provision.

We now show that excessive incentive provision and excessive benchmarking in private contracts give rise to crowded trades and excessive costs.

Proposition 3 (**Crowded Trades and Excessive Costs of Asset Management**)**.** *Compared to the equilibrium with privately optimal contracts, in*

the equilibrium with socially optimal contracts
(i) the risky asset's price is lower, $p^{**} < p^*$;
(ii) the fund's risky asset holdings are lower, $x^{M**} < x^{M*}$, *and the managers' costs are lower,* $\psi x^{M**} < \psi x^{M*}$.

As Proposition 3 shows, excessive use of incentive contracts by fund investors inflates the risky asset's price and reduces its expected return per share. In addition, the managers invest too much in the stock market and the costs of asset management are excessively high. Our model thus contributes to the debate on whether costs of asset management are excessive and whether returns delivered by the managers justify these costs.

We close the analysis with a few final remarks about the model.

Remark 1 (**Distributive Effects**)**.** Through our choice of weights in the social welfare function, we have shut down the contracts' distributive effects and isolated the pecuniary externality that the planner desires to correct. For certain applications, such as those related to wealth inequality, however, it could be interesting to analyze the transfers from one set of agents to another that benchmarking generates. Allowing for redistribution changes outcomes depending on whether an agent is a (net) buyer or a (net) seller of the risky asset. Because benchmarking boosts the risky asset price, this benefits (net) sellers of the risky asset at the expense of (net) buyers. If the social planner favors investors who have high endowments of the asset and are planning to sell (e.g., the older generations), she has incentives to use more benchmarking in order to inflate its price to assist them, and vice versa if she favors net buyers (who are typically the younger generations).

Remark 2 (**Prices Relative to the First Best**)**.** According to Proposition 3, $p^{**} < p^*$. We usually think of a constrained planner as being better at providing incentives than decentralized agents, and thus being closer to what an unconstrained planner can achieve. Surprisingly, the price of the risky asset in the first-best case exceeds equilibrium prices under both privately and socially optimal contracts, that is, $p^{**} < p^* < p^{FB}$.[37] So, the equilibrium price

[37]The expression for the first-best asset price is given in Lemma 5 in Appendix A. Com-

in the constrained optimum is not closer to the unconstrained-optimum price than the decentralized-equilibrium one, but are instead further away.[38]

While this might be surprising at first glance, this result is in fact quite intuitive. Under the first best, the portfolio is observable and it is optimal to choose high-alpha portfolios. This, of course, will push up the stock price and reduce its expected return. But, crowded trades are not a problem per se, because a pecuniary externality does not lead to an inefficiency in this case. In contrast, when the contract needs to provide incentives because the portfolio cannot be observed, a pecuniary externality does lead to an inefficiency, and crowded trades pose a problem as they reduce the effectiveness of incentive provision. While the comparison to the first best is irrelevant for practical purposes (because the first best is unattainable), it is helpful to highlight how exactly the mechanism that we explore works.

Remark 3 (**Achieving Social Optimum with Taxes**)**.** Given that privately optimal contracts result in an externality, it is natural to ask whether some sort of taxes could implement the constrained social optimum. We provide a detailed analysis of this question in Online Appendix D. We find the following. First, the manager's compensation needs to be (proportionally) taxed to make it more costly for the fund investor to provide incentives to the manager. This type of tax mimics the higher cost of incentive provision for the planner, who internalizes the externality. Second, the fund return net of the manager's compensation—which is the same as the fund investor's earnings in our model—should be (proportionally) subsidized. While this might be counterintuitive, the subsidy motivates the fund investor to lower a by increasing the benefit of keeping a larger $1-a$. An alternative to the subsidy is imposing a cap $\bar{a}$ on the fund manager's "skin in the game." Of course, the specific levels for the tax and subsidy rates (or, alternatively, the tax rate and the cap on a)

paring it to p^* given in Lemma 2 immediately yields the result.

[38]This result parallels that in Lorenzoni (2008), where the decentralized equilibrium falls between the constrained and unconstrained optima in terms of amount of borrowing and asset prices. However, in Lorenzoni's model the inequality signs in the price comparison are reverse—decentralized-equilibrium asset prices are lower than in the constrained optimum (higher in our model) and higher than in the first best (lower in our model).

depend on the model parameters (see Online Appendix D for the formulas).

Remark 4 (**Endogenizing the Choice of Becoming a Fund vs. Direct Investor**)**.** To zero in on the main mechanism we consider in the paper, we exogenously fixed the fractions of different agents in the population. One could endogenize the choice of becoming a fund investor or a direct investor, for example, by assuming a heterogeneous cost of participating in the asset market. This type of extension would introduce another channel through which crowded trades matter. The choices of individual investors of whether to be a fund investor or a direct investor, in the aggregate, would determine the size of the asset management sector. This in turn would affect the strength of the externality that we identify in the paper (i.e., how much contracts affect the risky asset's price and thus effectiveness of contracts designed by others). When making their decisions, the individual agents ignore this effect while the planner would account for this "extensive margin" of the externality when designing contracts.

Remark 5 (**Multiple Risky Assets**)**.** A natural extension of our model is to allow for multiple risky assets. In that case, the fund investors and the planner choose benchmark portfolio weights as part of the contract. Our main results fully extend to the multi-asset case. Moreover, the benchmark portfolio weights in privately and socially optimal contracts also differ (see Kashyap et al., 2022).

5 Conclusions

We consider the problem of optimal incentive provision for portfolio managers in a general-equilibrium asset-pricing model. The optimal contacts involve benchmarking. We show that by ignoring the effects of contracts on the equilibrium stock price, fund investors impose an externality on each other—the effectiveness of their incentive contracts is lower than they perceive them to be. Benchmarking boosts the stock price and lowers the expected return, making the marginal benefit of alpha-production lower for everyone. The social planner, who internalizes the effects of contracts on the equilibrium price, opts for less incentive provision, less benchmarking, and lower asset management costs.

In future work, it would be interesting to incorporate passive asset managers into the model. This extension is, however, not straightforward. The existing evidence on the compensation contracts in the asset management industry covers only active funds. Very little is known about contracts of managers in passive funds. Before engaging in modeling of passive managers, it would be important to collect such evidence. A natural starting point would be to analyze the Statements of Additional Information filed by the U.S. mutual funds with the Securities and Exchange Commission, which contain information on managers' compensation structure. If contracts of passive managers turn out to be incentive contracts, it would be interesting to understand the incentive problem they solve. It is not obvious what kind of incentive problem would result in optimal contracts that make the managers closely follow the benchmark. We leave this problem for future work.

Finally, environmental, social and governance (ESG) investing is one of the fastest growing segments in asset management. Another interesting extension would be to use this framework to study the optimal design of ESG benchmarks. We explore this problem in Kashyap et al. (2023*a*).

References

Acemoglu, Daron, and Alp Simsek. 2012. "Moral Hazard and Efficiency in General Equilibrium with Anonymous Trading." MIT working paper.

Admati, Anat R., and Paul Pfleiderer. 1997. "Does It All Add Up? Benchmarks and the Compensation of Active Portfolio Managers." *Journal of Business*, 70(3): 323–350.

Albuquerque, Rui A., Luis M. B. Cabral, and Jose Guedes. 2019. "Incentive Pay and Systemic Risk." *Review of Financial Studies*, 32: 4304–4342.

Anand, Amber, Chotibhak Jotikasthira, and Kumar Venkataraman. 2018. "Do Buy-side Institutions Supply Liquidity in Bond Markets? Evidence from Mutual Funds." *SMU Cox School of Business Research Paper.*

Basak, Suleyman, and Anna Pavlova. 2013. "Asset Prices and Institutional Investors." *American Economic Review*, 103(5): 1728–1758.

Berk, Jonathan B., and Richard C. Green. 2004. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy*, 112(6): 1269–1295.

Biais, Bruno, Florian Heider, and Marie Hoerova. 2021. "Variation Margins, Fire Sales, and Information-Constrained Optimality." *Review of Economic Studies*, 88(6): 2654–2686.

Brennan, Michael. 1993. "Agency and Asset Pricing." UCLA Working Paper.

Buffa, Andrea, Dimitri Vayanos, and Paul Woolley. 2014. "Asset Management Contracts and Equilibrium Prices." NBER Working Paper 20480.

Buffa, Andrea M., Dimitri Vayanos, and Paul Woolley. 2022. "Asset Management Contracts and Equilibrium Prices." *Journal of Political Economy*, 130(12): 3146–3201.

Calomiris, Charles W., Mauricio Larrain, Sergio L. Schmukler, and Tomas Williams. 2022. "Large International Corporate Bonds: Investor Behavior and Firm Responses." *Journal of International Economics*, 137: 103624.

Cuoco, Domenico, and Ron Kaniel. 2011. "Equilibrium Prices in the Presence of Delegated Portfolio Management." *Journal of Financial Economics*, 101(2): 264–296.

Cvitanic, Jaksa, and Hao Xing. 2018. "Asset Pricing Under Optimal Contracts." *Journal of Economic Theory*, 173(C): 142–180.

Davila, Eduardo, and Anton Korinek. 2018. "Pecuniary Externalities in Economies with Financial Frictions." *Review of Economic Studies*, 85: 352–395.

Di Tella, Sebastian. 2017. "Uncertainty Shocks and Balance Sheet Recessions." *Journal of Political Economy*, 125(6): 2038–2081.

Di Tella, Sebastian. 2019. "Optimal Regulation of Financial Intermediaries." *American Economic Review*, 109(1): 271–313.

Donaldson, Jason Roderick, and Giorgia Piacentino. 2018. "Contracting to Compete for Flows." *Journal of Economic Theory*, 173: 289–319.

Dybvig, Philip H., Heber K. Farnsworth, and Jennifer N. Carpenter. 2010. "Portfolio Performance and Agency." *The Review of Financial Studies*, 23(1): 1–23.

Eisele, Alexander, Tamara Nefedova, Gianpaolo Parise, and Kim Peijnenburg. 2020. "Trading Out of Sight: An Analysis of Cross-Trading in Mutual Fund Families." *Journal of Financial Economics*, 135: 359–378.

Fama, Eugene F, and Kenneth R French. 2010. "Luck versus Skill in the Cross-Section of Mutual Fund Returns." *Journal of Finance*, 65(5): 1915–1947.

Geanakoplos, John, and Heraklis M. Polemarchakis. 1996. "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete." in Gerard Debreu, ed., *General Equilibrium Theory, Vol. 2*, Cheltenham, UK: Elgar.

Greenwald, Bruce C., and Joseph E. Stiglitz. 1986. "Externalities in Economies with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics*, 90: 229–264.

Gromb, Denis, and Dimitri Vayanos. 2002. "Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs." *Journal of Financial Economics*, 66(2-3): 361–407.

Hart, Oliver D. 1975. "On the Optimality of Equilibrium When the Market Structure is Incomplete." *Journal of Economic Theory*, 11: 418–443.

He, Zhiguo, and Peter Kondor. 2016. "Inefficient Investment Waves." *Econometrica*, 84(2): 735–780.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics*, 10(1): 74–91.

Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica*, 55(2): 303–328.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization*, 7(Special Issue): 24–52.

Huang, Shiyang, Zhigang Qiu, and Liyan Yang. 2020. "Institutionalization, Delegation, and Asset Prices." *Journal of Economic Theory*, 186: 104977.

Ibert, Markus, Ron Kaniel, Stijn Van Nieuwerburgh, and Roine Vestman. 2017. "Are Mutual Fund Managers Paid for Investment Skill?" *Review of Financial Studies*, 31(2): 715–772.

Johnson, Travis L., and Gregory Weitzner. 2019. "Distortions Caused by Lending Fee Retention." University of Texas at Austin Working Paper.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng. 2008. "Unobserved Actions of Mutual Funds." *Review of Financial Studies*, 21(6): 2379–2416.

Kapur, Sandeep, and Allan Timmermann. 2005. "Relative Performance Evaluation Contracts and Asset Market Equilibrium." *The Economic Journal*, 115(506): 1077–1102.

Kashyap, Anil, Natalia Kovrijnykh, Jian Li, and Anna Pavlova. 2021. "The Benchmark Inclusion Subsidy." *Journal of Financial Economics*, 142(2): 756–774.

Kashyap, Anil, Natalia Kovrijnykh, Jian Li, and Anna Pavlova. 2022. "Is There Too Much Benchmarking in Asset Management?" NBER Working Paper No. 28020.

Kashyap, Anil, Natalia Kovrijnykh, Jian Li, and Anna Pavlova. 2023*a*. "Designing ESG Benchmarks." Work in Progress.

Kashyap, Anil, Natalia Kovrijnykh, Jian Li, and Anna Pavlova. 2023*b*. "Prices in Incentive Constraints." Work in Progress.

Keim, Donald B. 1999. "An Analysis of Mutual Fund Design: the Case of Investing in Small-Cap Stocks." *Journal of Financial Economics*, 51: 173–194.

Lines, Anton. 2016. "Do Institutional Incentives Distort Asset Prices?" Columbia Business School Working Paper.

Lorenzoni, Guido. 2008. "Inefficient Credit Booms." *Review of Economic Studies*, 75.

Ma, L., Yuehua Tang, and Juan-Pedro Gómez. 2019. "Portfolio Manager Compensation in the U.S. Mutual Fund Industry." *Journal of Finance*, 70: 587–638.

Ozdenoren, Emre, and Kathy Yuan. 2017. "Contractual Externalities and Systemic Risk." *Review of Economic Studies*, 84: 1789–1817.

Rinne, Kalle, and Matti Suominen. 2016. "Short-term Reversals, Returns to Liquidity Provision and the Costs of Immediacy." Aalto University Working Paper.

Sockin, Michael, and Mindy Z. Xiaolan. 2020. "Delegated Learning and Contract Commonality in Asset Management." University of Texas at Austin Working Paper.

Stoughton, Neal M. 1993. "Moral Hazard and the Portfolio Management Problem." *Journal of Finance*, 48(5): 2009–2028.

U.S. Securities and Exchange Commission. 2004. “Final Rule: Investment Adviser Codes of Ethics.” Release Nos. IA-2256, IC-26492; File No. S7-04-04.

A Proofs

Definition 1. *An equilibrium with privately optimal contracts is a contract* (a^*, b^*, c^*)*, the direct investor's portfolio* x^D*, the manager's portfolio* $x^M(a,b)$*, and a price* p^* *such that*
(i) given p^**,* x^D *solves the direct investor's problem* $\max_x x(\mu - p^*) - \gamma x^2\sigma^2/2$;
(ii) given p^**, for any* (a,b) $x^M(a,b)$ *solves the fund manager's problem*
$\max_x \; x(a\,\Delta - \psi) + (ax-b)(\mu - p^*) - \gamma\left[(ax-b)^2\sigma^2 + a^2\sigma_\varepsilon^2\right]/2$;
(iii) given p^* *and* $x^M(a,b)$*,* (a^*, b^*, c^*) *solves the fund investor's problem*

$$\max_{a,b,c} \; x(1-a)\Delta + [(1-a)x+b](\mu - p^*) - \frac{\gamma}{2}\left[[(1-a)x+b]^2\sigma^2 + (1-a)^2\sigma_\varepsilon^2\right] - c$$
$$s.t. \quad x(a\,\Delta - \psi) + (ax-b)(\mu - p^*) - \frac{\gamma}{2}\left[(ax-b)^2\sigma^2 + a^2\sigma_\varepsilon^2\right] + c \geq u_0,$$
$$x = x^M(a,b);$$

(iv) the stock market clears: $\lambda_D x^D + \lambda_M x^M(a^*, b^*) = \bar{x}$.

Throughout the paper, we use the following notation: $x^{M*} \equiv x^M(a^*, b^*)$.

Definition 2. *An equilibrium with socially optimal contracts is a contract* (a^{**}, b^{**}, c^{**})*, the direct investor's demand function* $x^D(p)$*, the manager's demand function* $x^M(p,a,b)$*, and a price function* $p = \hat{p}(a,b)$ *such that*
(i) for any p*,* $x^D(p)$ *solves the direct investor's problem* $\max_x x(\mu - p) - \gamma x^2\sigma^2/2$;
(ii) for any p *and* (a,b)*,* $x^M(p,a,b)$ *solves the fund manager's problem*
$\max_x x(a\,\Delta - \psi) + (ax-b)(\mu - p) - \gamma\left[(ax-b)^2\sigma^2 + a^2\sigma_\varepsilon^2\right]/2$;
(iii) given $\hat{p}(a,b)$*,* $x^D(p)$*, and* $x^M(p,a,b)$*,* (a^{**}, b^{**}, c^{**}) *solves the social planner's problem*

$$\max_{a,b,c} \quad \lambda_M\Big\{x_{-1}^F p + x(1-a)\Delta + [(1-a)x+b](\mu - p) - \frac{\gamma}{2}[(1-a)x+b]^2\sigma^2$$
$$- \frac{\gamma}{2}(1-a)^2\sigma_\varepsilon^2 - c\Big\} + \lambda_D\left\{x_{-1}^D p + x^D(p)(\mu - p) - \frac{\gamma}{2}\left[x^D(p)\right]^2\sigma^2\right\}$$
$$s.t. \; x(a\,\Delta - \psi) + (ax-b)(\mu - p) - \frac{\gamma}{2}\left[(ax-b)^2\sigma^2 + a^2\sigma_\varepsilon^2\right] + c \geq u_0,$$
$$x = x^M(p,a,b), \; p = \hat{p}(a,b);$$

(iv) the stock market clears: $\lambda_D x^D(\hat{p}(a^{**}, b^{**})) + \lambda_M x^M(\hat{p}(a^{**}, b^{**}), a^{**}, b^{**}) = \bar{x}$.

Throughout the paper, we use the following notation: $p^{**} \equiv \hat{p}(a^{**}, b^{**})$, $x^{M**} \equiv x^M(p^{**}, a^{**}, b^{**})$.

Proof of Lemma 1. (i) Equation (3) immediately follows from taking the first-order condition (FOC) of the direct investor's problem with respect to x. Similarly, (4) follows from taking the FOC of the manager's problem with respect to x. $a(\Delta - \psi/a + \mu - p) - \gamma\sigma^2(ax - b) = 0$. The second-order condition, $-\gamma a \sigma^2 < 0$, is (globally) satisfied so long as $a > 0$.

(ii) Substituting (3) and (4) in the market-clearing condition $\lambda_M x^M + \lambda_D x^D = \bar{x}$, we find the expression for the equilibrium asset price (5). □

Lemma 5 (**First Best**)**.** *If x is observable or if $\psi = 0$, then the optimal contract is $(a, b) = (1/2, 0)$, and the stock price is $p^{FB} = \mu - \gamma\sigma^2\bar{x} + 2\lambda_M(\Delta - \psi)$.*

Proof. When x is observable, the fund investor's problem is $\max_{a,b,x} x(\Delta - \psi + \mu - p) - \gamma\{(ax - b)^2\sigma^2 + [(1-a)x + b]^2\sigma^2 + [a^2 + (1-a)^2]\sigma_\varepsilon^2\}/2$. The FOC with respect to x is $x^M = (\Delta - \psi + \mu - p)/\{\gamma\sigma^2[a^2 + (1-a)^2]\} + (2a-1)b/[a^2 + (1-a)^2]$. The FOC with respect to b is $\gamma\sigma^2(y - z) = 0$, where $y = ax - b$ and $z = (1-a)x + b$. The FOC with respect to a is $-\gamma\sigma^2(y - z)x + \gamma(1 - 2a)\sigma_\varepsilon^2 = 0$, which, using the FOC with respect to b, implies $a = 1/2$. Then setting $b = 0$ satisfies the FOC with respect to b.

The portfolio choice evaluated at the optimal contract is $x^M = 2(\Delta - \psi + \mu - p)/(\gamma\sigma^2)$. Using market clearing, $p^{FB} = \mu - \gamma\sigma^2\bar{x} + 2\lambda_M(\Delta - \psi)$. Comparing with (14), $p^{FB} > p^*$. Finally, substituting $p = p^{FB}$ in x^M, the first-best portfolio of the manager is $x^M_{FB} = 2[\bar{x} + \lambda_D(\Delta - \psi)/(\gamma\sigma^2)]$.

Lastly, notice that if $\sigma_\varepsilon^2 = 0$, then the FOC with respect to b implies that the FOC with respect to a holds automatically. Thus, a and b are not separately pinned down. In particular, both $(a, b) = (1/2, 0)$ and $(a, b) = (1, \bar{x} + \lambda_D(\Delta - \psi)/(\gamma\sigma^2))$ are optimal. □

Proof of Lemma 2. (i) Using (6) and (8) and rearranging terms, (9) can be rewritten as $\gamma\sigma^2 b/a = (2 - 1/a)(\Delta - \psi + \mu - p) + (1 - 1/a)(1 - 1/a)\psi$. Using

(5), this implies

$$\frac{\gamma\sigma^2 b}{a} = \left(2 - \frac{1}{a}\right)\left[\Delta - \psi + \gamma\sigma^2\Lambda\left(\bar{x} - \lambda_M \frac{b}{a}\right) - \frac{\lambda_M \Lambda}{a}\left(\Delta - \frac{\psi}{a}\right)\right] + \left(1 - \frac{1}{a}\right)\left(\psi - \frac{\psi}{a}\right).$$

Rearranging terms, we have

$$\gamma\sigma^2 \underbrace{\left[1 + \left(2 - \frac{1}{a}\right)\lambda_M \Lambda\right]}_{=\Lambda} \frac{b}{a}$$
$$= \gamma\sigma^2\Lambda\left(2 - \frac{1}{a}\right)\bar{x} + \left(2 - \frac{1}{a}\right)\lambda_D \Lambda(\Delta - \psi) + \left[1 - \frac{1}{a} - \left(2 - \frac{1}{a}\right)\frac{\lambda_M \Lambda}{a}\right]\left(\psi - \frac{\psi}{a}\right),$$
$$\gamma\sigma^2\Lambda\frac{b}{a} = \Lambda\left(2 - \frac{1}{a}\right)\left[\gamma\sigma^2\bar{x} + \lambda_D(\Delta - \psi)\right] + \left[\frac{\lambda_M}{a} + \lambda_D - \frac{1}{a}\right]\Lambda\left(\psi - \frac{\psi}{a}\right),$$

implying (13). Equation (12) was derived in the main text.

(ii) Substituting (10) in the market clearing and rearranging terms yields (14). Substituting (10) in (4) and rearranging terms, implies $\gamma\sigma^2 x^{M*} = (\Delta - \psi + \mu - p^*) + (\Delta - \psi/a^* + \mu - p^*)$. Substituting (14) yields (15). □

Proof of Proposition 1. (i) Equation (12) has a negative and a positive root. We rule out the negative root because with $a < 0$ the manager's second-order condition is violated (and hence the first-order approach of writing the fund investor's problem is not valid)—see the proof of Lemma 1 in Online Appendix B. The right-hand side of (12) is strictly decreasing in a. It is strictly positive at $a = 1/2$ and strictly negative at $a = 1$. Thus $a^* \in (1/2, 1)$.

(ii) Suppose that $\bar{x} + \lambda_D(\Delta - \psi)/(\gamma\sigma^2) > 0$. Then $b^* > 0$ follows from (13) and part (i) of this proposition. □

Proof of Lemma 3. Rewrite (12) as $(2a^* - 1)a^{*3}\gamma^2\sigma^2/(1 - a^*) = \psi^2/\sigma_\varepsilon^2$. The left-hand side is increasing in a^*, while the right-hand side is increasing in ψ and decreasing in σ_ε^2. Thus a^* is increasing in ψ and decreasing in σ_ε^2. Moreover, rewriting the above equation as $(2a^* - 1)a^*\gamma^2\sigma^2\sigma_\varepsilon^2/(1 - a^*) = (\psi/a^*)^2$ we can see that the left-hand side is still increasing in a^*. Since a^* is increasing in ψ, ψ/a^* is increasing in ψ. The dependence of p^* and x^{M*} on ψ and σ_ε^2 then follows from (14) and (15). Moreover, a^* does not depend on Δ. Then the dependence of b^*, p^*, and x^{M*} on Δ follows from (13), (14), and (15). $\square$

Proof of Lemma 4. (i) Substituting the expression for z into (18) and rearranging terms, we have

$$0 = \Delta - \psi + \mu - p - \gamma\sigma^2 \left[\frac{\Delta - \psi/a + \mu - p}{\gamma\sigma^2} \left(\frac{1}{a} - 1 \right) + \frac{b}{a} \right] - \psi \frac{(1/a - 1)\Lambda\lambda_M}{1 - (1/a - 1)\Lambda\lambda_M}.$$

Rearranging terms,

$$\begin{aligned}
\gamma\sigma^2 \frac{b}{a} &= \Delta - \psi + \mu - p + \left(1 - \frac{1}{a}\right)\left(\Delta - \frac{\psi}{a} + \mu - p\right) - \psi \frac{(1-a)/a\Lambda\lambda_M}{1 - (1/a - 1)\Lambda\lambda_M}, \\
\gamma\sigma^2 b &= (2a-1)\left(\Delta - \psi + \mu - p\right) + (1-a)\left[\frac{1-a}{a} - \frac{\Lambda\lambda_M}{1 - (1/a-1)\Lambda\lambda_M}\right]\psi, \\
\gamma\sigma^2 b &= (2a-1)\left(\Delta - \psi + \mu - p\right) + (1-a)\left(\frac{1}{a} - \frac{1}{\lambda_M + \lambda_D}\right)\psi.
\end{aligned}$$

Substituting the expression for the price and rearranging terms, we have

$$b = (2a-1)\left[\bar{x} + \frac{1}{\gamma\sigma^2}\lambda_D(\Delta - \psi)\right] + (1-a)\left[\frac{1}{a} - \frac{\lambda_M/a + \lambda_D}{\lambda_M + \lambda_D}\right]\frac{\psi}{\gamma\sigma^2}.$$

Turning to the FOC with respect to a, use (17) to rewrite (21) as follows:

$$0 = \frac{1-a}{a}\frac{\lambda_M/a + \lambda_D}{\lambda_M + \lambda_D}\psi \left[\frac{\partial y}{\partial a} + \frac{\partial y}{\partial p}\frac{\partial p}{\partial a} + \frac{y}{a^2}\frac{\partial y}{\partial p}\frac{\partial p}{\partial (b/a)}\right] - (2a-1)\gamma\sigma_\varepsilon^2. \quad \text{(A.1)}$$

To express the term in square brackets, differentiate the market clearing $\lambda_M(y/a + b/a) + \lambda_D x^D = 0$ with respect to b/a and a and use $\partial x^D/\partial p = \partial y/\partial p$ to get

$$\begin{aligned}
\left(\frac{\lambda_M}{a} + \lambda_D\right)\frac{\partial y}{\partial p}\frac{\partial p}{\partial (b/a)} + \lambda_M = 0, \\
\left(\frac{\lambda_M}{a} + \lambda_D\right)\frac{\partial y}{\partial p}\frac{\partial p}{\partial a} - \lambda_M\frac{y}{a^2} + \frac{\lambda_M}{a}\frac{\partial y}{\partial a} = 0, \\
\left(\frac{\lambda_M}{a} + \lambda_D\right)\left[\frac{\partial y}{\partial p}\frac{\partial p}{\partial a} + \frac{y}{a^2}\frac{\partial y}{\partial p}\frac{\partial p}{\partial (b/a)}\right] + \frac{\lambda_M}{a}\frac{\partial y}{\partial a} = 0.
\end{aligned}$$

Then (A.1) becomes

$$0 = \frac{1-a}{a}\frac{\lambda_M/a+\lambda_D}{\lambda_M+\lambda_D}\psi\left[\frac{\partial y}{\partial a} - \frac{\partial y}{\partial a}\frac{\lambda_M/a}{\lambda_M/a+\lambda_D}\right] - (2a-1)\gamma\sigma_\varepsilon^2.$$

Finally, using $\partial y/\partial a = \psi/(\gamma\sigma^2 a^2)$, we obtain (23).

(ii) Substituting (20) in the market-clearing condition and rearranging terms yields (25). Substituting (20) in (4) and using (25) yields (26). □

Lemma 6. *The following inequality holds:*

$$\frac{1-a^*}{a^*}\left[\frac{1}{a^*} - \left(\frac{\lambda_M}{a^*}+\lambda_D\right)\right] > \frac{1-a^{**}}{a^{**}}\left[\frac{1}{a^{**}} - \frac{\lambda_M/a^{**}+\lambda_D}{\lambda_M+\lambda_D}\right].$$

Proof. See Online Appendix B. □

Proof of Proposition 2. (i) Comparison $a^{**} < a^*$ follows from comparing (12) and (23) and selecting the positive roots of the two equations, see the proof of Proposition 1 (i).

(ii) Denote $a_1 = a^*$, $a_2 = a^{**}$, $b_1 = b^*$, $b_2 = b^{**}$. From (13) and (24),

$$\frac{b_1}{a_1} = \left(2-\frac{1}{a_1}\right)\left[\bar{x} + \frac{1}{\gamma\sigma^2}\lambda_D(\Delta-\psi)\right] + \left(\frac{1}{a_1}-1\right)\left[\frac{1}{a_1} - \left(\frac{\lambda_M}{a_1}+\lambda_D\right)\right]\frac{1}{\gamma\sigma^2}\psi,$$
$$\frac{b_2}{a_2} = \left(2-\frac{1}{a_2}\right)\left[\bar{x} + \frac{1}{\gamma\sigma^2}\lambda_D(\Delta-\psi)\right] + \left(\frac{1}{a_2}-1\right)\left[\frac{1}{a_2} - \frac{\lambda_M/a_2+\lambda_D}{\lambda_M+\lambda_D}\right]\frac{1}{\gamma\sigma^2}\psi.$$

Under assumption $\bar{x}+\lambda_D(\Delta-\psi)/(\gamma\sigma^2) > 0$ and the fact that $a_1 > a_2$, Lemma 6 implies $b_1/a_1 > b_2/a_2$. Then using $a_1 > a_2 (> 1/2)$, $b_1 > b_2$ follows. □

Lemma 7. *The fund investor's and planner's second-order conditions are satisfied in equilibria with privately and socially optimal contracts, respectively.*

Proof. See Online Appendix B. □

Is There Too Much Benchmarking in Asset Management?

Anil Kashyap, Natalia Kovrijnykh, Jian Li, and Anna Pavlova

Online Appendices

B Omitted Proofs

The Fund Investor's Problem in Terms of Exponential Utilities:

$$\max_{a,b,c} - E\exp\left\{-\gamma\left[x^F_{-1}p + r_x - (ar_x - br_{\mathbf{b}}) - c\right]\right\}$$

subject to the manager's incentive constraint (4) and her participation constraint

$$-E\exp\left\{-\gamma\left[ar_x - br_{\mathbf{b}} + c\right]\right\} \geq \hat{u}_0, \tag{B.1}$$

where $\hat{u}_0$ is the exponential-utility version of u_0.[39] It is well known that in settings with normally distributed returns, CARA utility can be rewritten in a mean-variance form, leading to the problem described in subsection 4.2 in the main text.

The Social Planner's Problem in Terms of Exponential Utilities:

$$\max_{a,b,c} - \tilde{\omega}_F E\exp\left\{-\gamma\left[x^F_{-1}p + r_x - (ar_x - br_{\mathbf{b}}) - c\right]\right\}$$
$$- \tilde{\omega}_D E\exp\left\{-\gamma\left[x^D_{-1}p + x^D(D - p)\right]\right\}$$

subject to (3), (4), (5), and (B.1), where $\tilde{\omega}_i$, $i = F, F$, are the modified Pareto weights.

From the FOC with respect to c it follows that the Lagrange multiplier on the participation constraint equals $\tilde{\omega}_F MU_F/MU_M$, where MU_i denotes the expected marginal utility of agent i. This value is the effective Pareto weight on the manager's utility given that the contract allows transfer between the fund investor and manager (through c). Similarly, if transfers between fund and direct investors were allowed, then $\tilde{\omega}_F MU_F/\lambda_M = \tilde{\omega}_D MU_D/\lambda_D$,

[39]In particular, if the manager's outside option is risk-free, then $\hat{u}_0 = -\exp(-\gamma u_0)$.

and the distribution effects is zero. Without transfers, the Pareto weights that cancel out the distribution effects (in the formulation with exponential utilities) are equal to inverse marginal utilities times the population weights, $\tilde{\omega}_F = \lambda_M/MU_F$ and $\tilde{\omega}_D = \lambda_D/MU_D$.

Rewriting the objective function and the participation constraint in the mean-variance form gives the problem described in subsection 4.4 in the main text.

Lemma 6. The following inequality holds:

$$\frac{1-a^*}{a^*}\left[\frac{1}{a^*}-\left(\frac{\lambda_M}{a^*}+\lambda_D\right)\right] > \frac{1-a^{**}}{a^{**}}\left[\frac{1}{a^{**}}-\frac{\lambda_M/a^{**}+\lambda_D}{\lambda_M+\lambda_D}\right].$$

Proof. For expositional convenience, denote $a_1 = a^*$ and $a_2 = a^{**}$. Given that both sides of the above inequality are positive, it is equivalent to

$$\frac{(1-a_1)/a_1^2}{(1-a_2)/a_2^2}\frac{\lambda_M + a_1\lambda_D + (1-2a_1)\lambda_D}{(\lambda_M + a_2\lambda_D)\lambda_D/(\lambda_M+\lambda_D) + (1-2a_2)\lambda_D} > 1. \tag{B.2}$$

From (12) and (23) we have

$$\frac{1-a_1}{a_1^3(2a_1-1)} = \frac{1-a_2}{a_2^3(2a_2-1)}\frac{\lambda_D}{\lambda_M+\lambda_D}. \tag{B.3}$$

Substituting this in (B.2), obtain

$$\frac{a_1(2a_1-1)}{a_2(2a_2-1)}\frac{\lambda_D}{\lambda_M+\lambda_D}\frac{\lambda_M + a_1\lambda_D + (1-2a_1)\lambda_D}{(\lambda_M + a_2\lambda_D)\lambda_D/(\lambda_M+\lambda_D) + (1-2a_2)\lambda_D} > 1.$$

Since $a_1 > a_2$, it suffices to show that

$$\frac{\lambda_M + a_1\lambda_D + (1-2a_1)\lambda_D}{(\lambda_M + a_2\lambda_D)\lambda_D/(\lambda_M+\lambda_D) + (1-2a_2)\lambda_D} > \frac{\lambda_D+\lambda_M}{\lambda_D},$$

which is equivalent to

$$\frac{\lambda_M(2a_2-1)}{\lambda_D(a_1-a_2)} > 1. \tag{B.4}$$

To show (B.4), we will use equation (B.3). Rearranging (B.3) yields

$$\frac{1-a_1}{a_1^3(2a_1-1)}\frac{\lambda_M}{\lambda_D} = \frac{1-a_2}{a_2^3(2a_2-1)} - \frac{1-a_1}{a_1^3(2a_1-1)},$$

or, equivalently,

$$\frac{\lambda_M(2a_2-1)}{\lambda_D} = \frac{a_1^3}{1-a_1}\left[\frac{(1-a_2)(2a_1-1)}{a_2^3} - \frac{(1-a_1)(2a_2-1)}{a_1^3}\right].$$

The right-hand side of the above equation equals

$$\begin{aligned}&\frac{-a_1^3+2a_1^4-2a_1^4a_2+a_2a_1^3-(-a_2^3+2a_2^4-2a_2^4a_1+a_1a_2^3)}{(1-a_1)a_2^3}\\ &=\frac{(a_1-a_2)}{(1-a_1)a_2^3}\left[-(1+2a_1a_2)(a_1^2+a_1a_2+a_2^2)+2(a_1+a_2)(a_1^2+a_2^2)+a_1a_2(a_1+a_2)\right].\end{aligned}$$

Rearranging terms and doing some more algebra, yields

$$\frac{\lambda_M(2a_2-1)}{\lambda_D(a_1-a_2)} = \frac{(2a_1-1)a_1^2(1-a_2)+(2a_2-1)a_2^2(1-a_1)+(2a_1-1)a_1a_2+2a_1a_2^2(1-a_1)}{a_2^3(1-a_1)}.$$

Since $1/2 < a_2 < a_1 < 1$,

$$\frac{\lambda_M(2a_2-1)}{\lambda_D(a_1-a_2)} > \frac{(2a_1-1)a_1^2(1-a_2)+(2a_2-1)a_2^2(1-a_1)+(2a_1-1)a_1a_2+a_2^3(1-a_1)}{a_2^3(1-a_1)} > 1,$$

and thus (B.4) holds. □

Lemma 7. The fund investor's and social planner's second-order conditions are satisfied in the equilibria with privately and socially optimal contracts, respectively.

Proof of Lemma 7. Denote by $F_{b/a}$ and F_a the left-hand sides of the FOCs with respect to b/a and a, respectively. From the proofs of Lemmas 2 and 4, once we plug in the FOC with respect to b/a in the FOC with respect to a, the remaining terms only depend a. Thus we can write F_a in the following form: $F_a = g(a) + F_{b/a}h(a, b/a)$. The function $g(a)$

is given by (the right-hand sides of) equations (12) and (23) with privately and socially optimal contracts, respectively.

Differentiating F_a with respect to a and b/a,

$$F_{aa} = \frac{\partial F_a}{\partial a} = g'(a) + \underbrace{F_{b/a}}_{=0} \frac{\partial h(a, b/a)}{\partial a} + F_{b/a,a} h(a, b/a),$$

$$F_{a,b/a} = \frac{\partial F_a}{\partial (b/a)} = \underbrace{F_{b/a}}_{=0} \frac{\partial h(a, b/a)}{\partial (b/a)} + F_{b/a,b/a} h(a, b/a).$$

Notice that $g'(a) < 0$ (this follows from (12) with privately optimal contracts and from (23) with socially optimal contracts). Furthermore, $F_{b/a,b/a} < 0$. Indeed, in the privately optimal case, $F_{b/a,b/a} = -\gamma\sigma^2/a < 0$. Similarly, in the socially optimal case, $F_{b/a,b/a} = -\gamma\Lambda\lambda_M\sigma^2 - \gamma\sigma^2/a < 0$. Finally,

$$\det \begin{pmatrix} F_{b/a,b/a} & F_{a,b/a} \\ F_{a,b/a} & F_{a,a} \end{pmatrix} = \det \begin{pmatrix} F_{b/a,b/a} & F_{b/a,b/a} h \\ F_{b/a,b/a} h & g'(a) + F_{b/a,b/a} h^2 \end{pmatrix} = g'(a) F_{b/a,b/a} > 0.$$

This completes the proof. □

C Discussion on Value Added and Costs of Asset Management

This appendix elaborates on the assumptions we make regarding the costs and benefits of asset management.

As mentioned in the body of the paper, there are a variety of interpretations for alpha. In our formulation, alpha has nothing to do with superior information, which could be associated with stock-selection and market-timing abilities. Under this interpretation, direct investors who happen to buy the same assets or traded at the same time still do not earn the same returns as the managers. This interpretation has the advantage of being consistent with the vast literature (e.g., Fama and French, 2010) that casts doubt on the ability to generate abnormal returns by stock picking or market timing.

It is also consistent with a great deal of empirical evidence suggesting that savvy investors can augment their returns by lending securities, by conserving on transactions costs (e.g., from crossing trades in-house or by obtaining favorable quotes from brokers) or by providing liquidity (i.e., serving as a counterparty to liquidity demanders and earning a premium on such trades). For example, securities lending contributed 5% of total revenue of both BlackRock and State Street in 2017. While it has recently become possible for some retail investors to participate in securities lending, they earn lower returns for this activity and do not have the same opportunities as a large asset management firm. It is also well established that portfolio managers can profit from providing immediacy in trades, by either buying assets which are out of favor or selling ones that are in high demand.[40] It would be prohibitively expensive for retail investors to try to do this. Finally, Eisele et al. (2020) present evidence that trades crossed internally within a fund complex are executed more cheaply than comparable external trades.

The noise term ε in (1) captures the fact that the return-augmenting activities do not produce a certain return each period. For example, the demand for liquidity, the opportunities to lend shares and the possibility of crossing trades all fluctuate, so even a very alert and skilled manager will have some randomness in her returns. Also for securities that are lent, there is a risk that they will not be returned in a timely manner or potentially at all.

There is also considerable evidence to support our assumption that the manager must incur a private cost in order to deliver the abnormal returns. For instance, to successfully buy and sell at the appropriate times to provide liquidity, the manager has to be actively monitoring market conditions while markets are open. For securities lending, the manager would also have to decide whether to accommodate requests to borrow shares. In some cases, these demands arise because the entity borrowing the shares wants to vote them and the manager must decide whether to pass up that choice.[41]

[40]In a classic paper, Keim (1999) estimates an annual alpha of 2.2% earned by liquidity provision activities of a fund. Rinne and Suominen (2016) document that the top decile of liquidity providing mutual funds outperform the bottom decile by about 60 basis points per year. Anand, Jotikasthira and Venkataraman (2018) find similar estimates using a different sample of funds over a different time period.

[41]Most managers also incur some costs that are observable and can be passed on directly to fund

We could instead assume that the private cost arises because the manager needs to exert costly effort to generate the excess returns, as is often done in the contracting literature (e.g., Holmstrom and Milgrom, 1987, 1991). Incorporating effort makes the algebra much more involved.[42] However, under certain assumptions our main insights extend to this case. Importantly, it is the unobservability of the portfolio holdings and not the unobservability of effort that is central to our mechanism. To make this clear and to focus on the key friction, in our main model we do not include an effort choice. We analyze an extension that incorporates effort in Appendix E.1 and show that our main insights carry over.

It is also plausible that the benefits and costs associated with the return-augmenting activities are increasing in the size of the holdings.[43] For example, in terms of the liquidity provision and trade-crossing, the wider the range of securities in the portfolio and/or the more a fund holds on any particular security, the easier it would be to provide liquidity or more likely it would be that a trade can be offset. For securities lending, a larger portfolio opens up additional lending opportunities. As mentioned earlier, it is simplest to think of the costs as being tied to the time it takes to undertake the various activities. Thought of this way, if the opportunities to augment returns increase as the portfolio expands, then the costs of realizing them would naturally grow too.

investors. Examples would include custody, audit, shareholder reports, proxies and some external legal fees. Our main results continue to hold in a model in which some costs are observable.

[42]Our results trivially extend if effort is bounded from above (e.g., if there is a time constraint), and the optimal solution is at the upper bound.

[43]Implicit in our expressions for the return on the fund in (1) and the portfolio-management cost is that they scale linearly with the size of the portfolio. This is seemingly inconsistent with Berk and Green (2004) who assume that there are decreasing returns to scale in asset management, but it is not. Berk and Green explicitly attribute decreasing returns to scale to the price impact of fund managers. The bigger the portfolio invested in an alpha-opportunity, the smaller the return on a marginal dollar invested. Berk and Green's model is in partial equilibrium and their price impact is simply an exogenous function of fund size. Ours is a general-equilibrium model, in which the price impact *endogenously* arises from a higher aggregate demand of portfolio managers for the risky asset. Linearity allows us to solve the model in closed form, but what is important conceptually is that the cost is increasing in x. We show in Appendix E.1 that while the algebra is messier, under some assumptions our main analysis extends to the case of more general specifications of the return and cost.

D Achieving the Social Optimum with Taxes

This appendix analyzes how imposing taxes can implement the constrained socially optimal allocation and stock price in the equilibrium in which contracts are chosen by fund investors. There are multiple ways of doing that, and we consider two alternatives here—one with proportional income taxes (or subsidies) on the managers and fund investors, the other with an income tax on the managers and a cap on a.[44]

First, suppose there are proportional tax rates on the fund investors' and managers' incomes, denoted by t and t', respectively. The tax revenue—which is uncertain, given that the incomes are uncertain—is distributed to the fund investors as a lump-sum transfer T. Denote the constant and stochastic part of the transfer by τ_0 and τ so that $T = \tau_0 + \tau(\tilde{D} - p)$. How τ_0 and τ are determined is discussed later.

Since we want to implement the constrained optimal allocation, the taxes and the lump-sum transfer will be such that $y = (1 - t')[ax - b]$ and $z = (1 - t)[(1 - a)x + b] + \tau$ are the same as in the constrained social optimum.

The utilities of the fund investor and manager with taxes can be written as

$$U^F = (1 - t)(1 - a)x\Delta + z(\mu - p) - c(1 - t) + \tau_0 - \frac{\gamma}{2}\left[z^2\sigma^2 + (1 - t)^2(1 - a)^2\sigma_\varepsilon^2\right] + x_{-1}^F p,$$
$$U^M = (1 - t')ax\Delta - x\psi + y(\mu - p) + c(1 - t') - \frac{\gamma}{2}\left[y^2\sigma^2 + (1 - t')^2 a^2 \sigma_\varepsilon^2\right].$$

The manager's demand function is

$$x^M = \frac{\Delta - \psi/[a(1 - t')] + \mu - p}{\gamma\sigma^2 a(1 - t')} + \frac{b(1 - t')}{a(1 - t')}. \tag{D.1}$$

To implement the social optimum, we need $a(1 - t') = a^{**}$ and $b(1 - t') = b^{**}$.

From the first-order condition with respect to c, the Lagrange multiplier on the man-

[44] As will become clear from the analysis, we need two tax rates to eliminate the differences in the two first-order conditions (with respect to b/a and a) in the private and social cases, and one tax rate is not enough.

ager's participation constraint is $\xi = (1-t)/(1-t')$. The fund investor maximizes

$$\begin{aligned} U^F + \xi U^M &= [(1-t)x + \tau](\Delta + \mu - p) + \tau_0 - \frac{1-t}{1-t'}x\psi \\ &- \frac{\gamma}{2}\left\{z^2\sigma^2 + \frac{1-t}{1-t'}y^2\sigma^2 + (1-t)\left[(1-t)(1-a)^2 + (1-t')a^2\right]\sigma_\varepsilon^2\right\} \end{aligned}$$

subject to the manager's incentive constraint (D.1), $y = (1-t')[ax - b]$, and

$$z = (1-t)\left[\frac{1}{1-t'}\frac{1-a}{a}y + \frac{b}{a}\right] + \tau.$$

The first-order condition with respect to b/a is

$$\begin{aligned} &(1-t)(\Delta + \mu - p - \gamma\sigma^2 z) - \frac{1-t}{1-t'}\psi = 0, \\ &\Delta + \mu - p - \gamma\sigma^2 z - \frac{1}{1-t'}\psi = 0. \end{aligned} \tag{D.2}$$

Recall that the planner's first-order condition with respect to b/a is

$$\Delta + \mu - p - \gamma\sigma^2 z - \psi\frac{\lambda_M/a^{**} + \lambda_D}{\lambda_M + \lambda_D} = 0.$$

To equate the two, we need $1 - t' = (\lambda_M + \lambda_D)(\lambda_M/a^{**} + \lambda_D)$, or

$$t' = \frac{\lambda_M}{\lambda_M/a^{**} + \lambda_D}\frac{1-a^{**}}{a^{**}}. \tag{D.3}$$

Intuitively, the positive tax on the manager's income inflates his costs relative to returns, which discourages him from investing in the risky asset.

The first-order condition with respect to a is

$$(1-t)\left[(1-t)(1-a) - (1-t')a\right]\gamma\sigma_\varepsilon^2 + (\Delta + \mu - p + \gamma\sigma^2 z)\frac{1-t}{1-t'}\frac{1-a}{a}\frac{\partial y}{\partial a} = 0.$$

Dividing by $1-t$ and using (D.2), $\partial y/\partial a = \psi/(\gamma\sigma^2 a^2(1-t'))$, and $a(1-t') = a^{**}$, the above condition can be rewritten as

$$[(1-t)(1-a)-(1-t')a]\,\gamma\sigma_\varepsilon^2+\frac{1-a}{a^{**3}}\frac{\psi^2}{\gamma\sigma^2}=0.$$

Recall that the planner's first-order condition with respect to a is

$$(1-2a^{**})\gamma\sigma_\varepsilon^2+\frac{1-a^{**}}{a^{**3}}\frac{\psi^2}{\gamma\sigma^2}\frac{\lambda_D}{\lambda_M+\lambda_D}=0.$$

To equate the two, we need

$$\frac{1-a}{(1-t)(1-a)-(1-t')a}=\frac{\lambda_D}{\lambda_M+\lambda_D}\frac{1-a^{**}}{1-2a^{**}}, \tag{D.4}$$

From $a=a^{**}/(1-t')=a^{**}(\lambda_M/a^{**}+\lambda_D)/(\lambda_M+\lambda_D)$, $1-a=(1-a^{**})\lambda_D/(\lambda_M+\lambda_D)$, and (D.4) simplifies to $(1-t)(1-a)-(1-t')a=1-2a^{**}$, or

$$t(1-a)+t'a=0. \tag{D.5}$$

Using the expression for t' given in (D.3) and $a=a^{**}/(1-t')$, we have

$$t=-\lambda_M/\lambda_D.$$

That is, in order to implement the constrained social optimum, the fund manager's income tax rate should be negative. Intuitively, in order to discourage the fund investor from setting a too high, the subsidy should be used so that the fund investor effectively retains a larger share of the return for himself. His after-tax share of the return equals $(1-t)(1-a)=1-(1-t')a$. That is, it is as if he only has to give $(1-t')a$ instead of a to the manager. Thus the income tax rates t and t' considered here effectively translate into the tax rates of t' imposed directly on a and b such that $(1-t')a=a^{**}$ and $(1-t')b=b^{**}$.

Finally, the transfer to the fund investor that balances the budget is

$$\begin{aligned} T &= [t(1-a) + t'a]x(\Delta + \tilde{D} - p) + (t - t')[b(\tilde{D} - p) - c] \\ &= (t - t')[b(\tilde{D} - p) - c], \end{aligned}$$

where the last equality follows from (D.5), and so $\tau_0 = (t - t')c$ and $\tau = (t - t')b$. Note that while $t - t' < 0$, the expected lump-sum transfer $(t - t)'\, [b(\mu - p) - c]$ can be negative or positive depending on the value of the manager's outside option, which pins down c.

An alternative scheme that achieves the social optimum is a combination of the income tax rate t' given by (D.3) imposed on the manager together with a cap (an upper bound) on the sensitivity of the manager's compensation with respect to the fund performance, a, at $\bar{a} = a^{**}/(1 - t')$, so that $a \leq \bar{a} = (\lambda_M + a^{**}\lambda_D)/(\lambda_M + \lambda_D)$. As before, the total amount of tax revenue should be paid to the fund investor as a lump-sum transfer.

E Extensions

E.1 Incorporating an Effort Choice by the Manager

In this appendix we extend the model in the main text to incorporate an effort choice by the manager. We will assume here that the effort choice is unobservable to the fund investor (the analysis of the case with observable effort is similar). We still assume, as in the main text, that the manager's portfolio choice is unobservable as well. We will demonstrate that our main insights extend in this case. In particular, the individual fund managers overestimate the effectiveness of incentive provision relative to the planner, which results in crowded trades.

Consider general functional forms so that the benefit function is $\tilde{\Delta}(x, e)$, the cost function is $\tilde{\psi}(x, e)$, and the variance of the noise term is $\tilde{\varepsilon}(x, e)$.

The manager's problem is

$$\max_{x,e}\ a\tilde{\Delta}(x, e) - \tilde{\psi}(x, e) + (ax - b)(\mu - p) - \frac{\gamma}{2}\sigma^2(ax - b)^2 - \frac{\gamma}{2}a^2\tilde{\varepsilon}(x, e) + c.$$

The first-order conditions with respect to e is

$$\frac{\partial \tilde{\Delta}}{\partial e} - \frac{1}{a}\frac{\partial \tilde{\psi}}{\partial e} - \frac{\gamma}{2}a\frac{\partial \tilde{\varepsilon}}{\partial e} = 0. \tag{E.1}$$

Think of the optimal effort solving (E.1) as $e^*(x, a)$.

We impose the following assumptions.

Assumption 1. *Suppose that for each $a \in [1/2, 1]$, the function*

$$a\tilde{\Delta}(x, e) - \tilde{\psi}(x, e) - \frac{\gamma a^2}{2}\left[x^2 + \varepsilon(x, e)\right]$$

is concave in (x, e). Moreover, denote

$$\frac{df(x, e^*(x, a))}{dx} = \frac{\partial f}{\partial e}\frac{\partial e^*}{\partial x} + \frac{\partial f}{\partial x},$$

where function f is either $\tilde{\Delta}$, $\tilde{\psi}$, or $\tilde{\varepsilon}$, and $e^*(x, a)$ is implicitly defined by (E.1). Suppose that for each $a \in [1/2, 1]$,

$$\frac{d\psi}{dx} > \frac{\gamma}{2}\left|\frac{d\varepsilon}{dx}\right|, \quad \frac{d^2\psi}{dx^2} \geq \frac{\gamma}{2}\left|-\frac{d^2\varepsilon}{dx^2}\right|.$$

The above inequalities require that the manager's private cost is sufficiently increasing and sufficiently convex in x (once the optimal effort choice is taken into account).

We now proceed with the analysis of the manager's problem. The manager's first-order condition with respect to x (taking into account the fact that x affects the optimal choice of effort according to $e^*(x, a)$) is

$$\mu - p - \gamma\sigma^2(ax - b) + \frac{d\tilde{\Delta}}{dx} - \frac{1}{a}\frac{d\tilde{\psi}}{dx} - \frac{\gamma}{2}a\frac{d\tilde{\varepsilon}}{dx} = 0. \tag{E.2}$$

Assumption 1 implies that the second-order conditions are satisfied, in particular,

$$SOC_x \equiv -\gamma\sigma^2 a + \frac{d^2\tilde{\Delta}}{dx^2} - \frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2} - \frac{\gamma a}{2}\frac{d^2\tilde{\varepsilon}}{dx^2} < 0.$$

In what follows, we will use expressions for the effects of b and a on x that we derive below. Differentiating (E.2) with respect to b,

$$\gamma\sigma^2 + SOC_x \frac{\partial x}{\partial b} = 0,$$
$$\frac{\partial x}{\partial b} = -\frac{\gamma\sigma^2}{SOC_x} = \frac{\gamma\sigma^2}{\gamma\sigma^2 a - \frac{d^2\tilde{\Delta}}{dx^2} + \frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2} + \frac{\gamma}{2}a\frac{d^2\tilde{\varepsilon}}{dx^2}} > 0.$$

Denote $\frac{dx}{di} \equiv \frac{\partial x}{\partial i} + \frac{\partial x}{\partial p}\frac{\partial p}{\partial i}$, $i \in \{a, b\}$. Taking the total derivative of (E.2) with respect to b,

$$\gamma\sigma^2 - \frac{\partial p}{\partial b} + SOC_x \frac{dx}{db} = 0. \tag{E.3}$$

Differentiating the market-clearing condition $\lambda_M x + \lambda_D x^D = \bar{x}$ with respect to b (and using the expression for x^D in the main text),

$$\lambda_M \frac{dx}{db} + \lambda_D \frac{\partial x^D}{\partial p}\frac{\partial p}{\partial b} = \lambda_M \frac{dx}{db} - \lambda_D \frac{1}{\gamma\sigma^2}\frac{\partial p}{\partial b} = 0,$$
$$\frac{\partial p}{\partial b} = \gamma\sigma^2 \frac{\lambda_M}{\lambda_D}\frac{dx}{db}.$$

Substituting this into (E.3), yields

$$\frac{dx}{db} = \frac{\gamma\sigma^2}{\gamma\sigma^2\frac{\lambda_M}{\lambda_D} - SOC_x} = \frac{\gamma\sigma^2}{\gamma\sigma^2\left(a + \frac{\lambda_M}{\lambda_D}\right) - \frac{d^2\tilde{\Delta}}{dx^2} + \frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2} + \frac{\gamma}{2}a\frac{d^2\tilde{\varepsilon}}{dx^2}}.$$

Notice that $\frac{dx}{db} \leq \frac{\partial x}{\partial b}$, with strict inequality if $\lambda_M > 0$.

Similarly, differentiating (E.2) with respect to a, gives

$$
\begin{aligned}
&-\gamma\sigma^2 x + \frac{1}{a^2}\frac{d\tilde{\psi}}{dx} - \frac{\gamma}{2}\frac{d\tilde{\varepsilon}}{dx} + SOC_x \frac{\partial x}{\partial a} = 0, \\
&\frac{\partial x}{\partial a} = \frac{1}{\gamma\sigma^2 a - \frac{d^2\tilde{\Delta}}{dx^2} + \frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2} + \frac{\gamma}{2}a\frac{d^2\tilde{\varepsilon}}{dx^2}}\left[\frac{1}{a^2}\frac{d\tilde{\psi}}{dx} - \frac{\gamma}{2}\frac{d\tilde{\varepsilon}}{dx}\right] - x\frac{\partial x}{\partial b}.
\end{aligned} \tag{E.4}
$$

The last term captures the negative effect of a on x because the manager is exposed to too much aggregate risk—the effect which b offsets. There is a new effect that we did not have before—a larger a reduces x if $\tilde{\varepsilon}$ is increasing in x because it exposes the manager to more idiosyncratic risk, and this risk cannot be offset by an increase in b. Notice that without it (as in the main text), we would have $\partial x/\partial a + x\partial x/\partial b > 0$, which captures the fact with b offsetting the negative effect of a on x, we are only left with the positive effect that is coming from reducing the effective cost. We want to make sure that $\partial x/\partial a + x\partial x/\partial b > 0$. Notice that if this was not the case, it would not be optimal for the fund investor to use a for incentive provision purposes. Assumption 1 ensures that, and we have

$$
\frac{\partial x}{\partial a} + x\frac{\partial x}{\partial b} = \frac{\frac{1}{a^2}\frac{d\tilde{\psi}}{dx} - \frac{\gamma}{2}\frac{d\tilde{\varepsilon}}{dx}}{\gamma\sigma^2 a - \frac{d^2\tilde{\Delta}}{dx^2} + \frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2} + \frac{\gamma}{2}a\frac{d^2\tilde{\varepsilon}}{dx^2}} > 0.
$$

Similarly, we have

$$
\frac{dx}{da} + x\frac{dx}{db} = \frac{\frac{1}{a^2}\frac{d\tilde{\psi}}{dx} - \frac{\gamma}{2}\frac{d\tilde{\varepsilon}}{dx}}{\gamma\sigma^2\left(a + \frac{\lambda_D}{\lambda_M}\right) - \frac{d^2\tilde{\Delta}}{dx^2} + \frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2} + \frac{\gamma}{2}a\frac{d^2\tilde{\varepsilon}}{dx^2}},
$$

which is smaller than $\partial x/\partial a + x\partial x/\partial b$.

We now turn to the analysis of the fund investor's problem. Denoting $y = ax - b$ and

$z = x - y$, this problem is

$$\max_{a,b,c,x} \quad (1-a)\tilde{\Delta}(x, e^*(x,a)) + z(\mu - p) - \frac{\gamma\sigma^2}{2}z^2 - \frac{\gamma(1-a)^2}{2}\tilde{\varepsilon}^2(x, e^*(x,a)) - c$$

subject to the manager's participation constraint and incentive constraint (E.2) (in which we substituted $e^*(x,a)$ implicitly defined by (E.1)).

The fund investor's first-order condition with respect to b is

$$\frac{d(U^F + U^M)}{db} = \frac{\partial U^F}{\partial x}\frac{\partial x}{\partial b} + \underbrace{\frac{\partial U^M}{\partial x}}_{=0}\frac{\partial x}{\partial b} + \frac{\partial(U^F + U^M)}{\partial b} = 0. \tag{E.5}$$

The last term captures how b directly affects the social welfare by linearly transferring from y to z. The first term captures the indirect effect of b on social welfare through its effect on the manager's demand x. Intuitively, notice that $\partial U^F/\partial x$ should be positive, otherwise b would not be positive. We will show that $\partial U^F/\partial x > 0$ formally below. The last term in (E.5) is

$$\frac{\partial(U^F + U^M)}{\partial b} = -\frac{\gamma\sigma^2}{2}\frac{\partial(y^2 + z^2)}{\partial b} = \gamma\sigma^2(y - z) = \gamma\sigma^2\left[(2a-1)x - 2b\right].$$

We will show below that this term is negative (notice that this term would be zero under perfect risk sharing $a = 1/2$ and $b = 0$.)

Using (E.2),

$$\begin{aligned}\frac{\partial U^F}{\partial x} &= (1-a)\left[\frac{d\tilde{\Delta}}{dx} + \mu - p - \gamma\sigma^2 z - \frac{\gamma}{2}(1-a)\frac{d\tilde{\varepsilon}}{dx}\right] \\ &= (1-a)\left[\gamma\sigma^2(y-z) + \frac{1}{a}\frac{d\tilde{\psi}}{dx} + \frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right].\end{aligned}$$

Then the investor's first-order condition with respect to b becomes

$$(1-a)\left[\gamma\sigma^2(y-z) + \frac{1}{a}\frac{d\tilde{\psi}}{dx} + \frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right]\frac{\partial x}{\partial b} + \gamma\sigma^2(y-z) = 0, \tag{E.6}$$

or equivalently

$$\frac{(1-a)\dfrac{\partial x}{\partial b}}{(1-a)\dfrac{\partial x}{\partial b}+1}\left[\frac{1}{a}\frac{d\tilde{\psi}}{dx}+\frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right]+\gamma\sigma^2(y-z)=0. \tag{E.7}$$

Notice that since the first term is strictly positive by Assumption 1, the second term is strictly negative. It then also follows that the term in the square brackets in E.6 must be strictly positive, that is, $\partial U^F/\partial x = \partial(U^F+U^M)/\partial x > 0$. Intuitively, it means that it is optimal for the fund investor to use contracts to provide incentives. It also then follows that $b > 0$. Indeed, notice that at $b = 0$ and $a \in [1/2, 1]$, the left-hand side of (E.7) is strictly positive given Assumption 1, and thus $b \leq 0$ cannot be optimal.

We will now compare the social planner's first-order condition with respect to b to that of an individual fund investor. The planner's first-order condition with respect to b (after canceling out the distributive effects, as in the main text) is the same as the corresponding first-order condition for an investor, but $\partial x/\partial b$ is being replaced with dx/db, namely

$$\frac{\partial U^F}{\partial x}\frac{dx}{db}+\frac{\partial(U^F+U^M)}{\partial b}=0,$$

or

$$\frac{(1-a)\dfrac{dx}{db}}{(1-a)\dfrac{dx}{db}+1}\left[\frac{1}{a}\frac{d\tilde{\psi}}{dx}+\frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right]+\gamma\sigma^2(y-z)=0.$$

Since $dx/db < \partial x/\partial b$ as long as $\lambda_M > 0$,

$$\frac{(1-a)\dfrac{dx}{db}}{(1-a)\dfrac{dx}{db}+1}<\frac{(1-a)\dfrac{\partial x}{\partial b}}{(1-a)\dfrac{\partial x}{\partial b}+1}.$$

It then follows that under Assumption 1, the additional terms in the planner's first-order condition relative to the investor's first-order condition are strictly negative.

Now consider the first-order condition with respect to a. In the privately optimal case, it is

$$\frac{d(U^F+U^M)}{da}=\frac{\partial U^F}{\partial x}\frac{\partial x}{\partial a}+\frac{\partial U^F}{\partial e}\frac{\partial e}{\partial a}+\frac{\partial (U^F+U^M)}{\partial a}=0.$$

Rewrite this to get

$$\begin{aligned}\frac{d(U^F+U^M)}{da}&=(1-a)\left[\gamma\sigma^2(y-z)+\frac{1}{a}\frac{d\tilde{\psi}}{dx}+\frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right]\frac{\partial x}{\partial a}\\&+(1-a)\left[\frac{\partial\tilde{\Delta}}{\partial e}-\frac{\gamma}{2}(1-a)\frac{\partial\tilde{\varepsilon}}{\partial e}\right]\frac{\partial e}{\partial a}-\gamma\sigma^2(y-z)x-\gamma\varepsilon^2(2a-1).\\&=(1-a)\left[\gamma\sigma^2(y-z)+\frac{1}{a}\frac{d\tilde{\psi}}{dx}+\frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right]\frac{\partial x}{\partial a}\\&+(1-a)\left(\frac{1}{a}\frac{\partial\tilde{\psi}}{\partial e}+\frac{\gamma}{2}(2a-1)\frac{\partial\tilde{\varepsilon}}{\partial e}\right)\frac{\partial e}{\partial a}-\gamma\sigma^2(y-z)x-\gamma\varepsilon^2(2a-1)=0.\end{aligned}$$

where the second equality uses (E.1). Then using (E.6), we can rewrite the above condition as follows:

$$\begin{aligned}&(1-a)\left[\gamma\sigma^2(y-z)+\frac{1}{a}\frac{d\tilde{\psi}}{dx}+\frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right]\left(\frac{\partial x}{\partial a}+x\frac{\partial x}{\partial b}\right)\\&+(1-a)\left(\frac{1}{a}\frac{\partial\tilde{\psi}}{\partial e}+\frac{\gamma}{2}(2a-1)\frac{\partial\tilde{\varepsilon}}{\partial e}\right)\frac{\partial e}{\partial a}-\gamma\varepsilon^2(2a-1)=0.\end{aligned}$$

Using (E.7), the fund investor's first-order condition with respect to a becomes

$$\begin{aligned}&\frac{(1-a)\left(\frac{\partial x}{\partial a}+x\frac{\partial x}{\partial b}\right)}{(1-a)\frac{\partial x}{\partial b}+1}\left[\frac{1}{a}\frac{d\tilde{\psi}}{dx}+\frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right]+(1-a)\left(\frac{1}{a}\frac{\partial\tilde{\psi}}{\partial e}+\frac{\gamma}{2}(2a-1)\frac{\partial\tilde{\varepsilon}}{\partial e}\right)\frac{\partial e}{\partial a}\\&-\gamma\varepsilon^2(2a-1)=0.\end{aligned}\tag{E.8}$$

Notice that we need $d\tilde{\psi}/dx>0$ or $\partial\tilde{\psi}/\partial e>0$, otherwise $a=1/2$ is optimal. This is guaranteed by Assumption 1.

The social planner's first-order condition with respect to a is obtained from (E.8) by replacing

$$\frac{(1-a)\left(\frac{\partial x}{\partial a}+x\frac{\partial x}{\partial b}\right)}{(1-a)\frac{\partial x}{\partial b}+1}=\frac{\left(\frac{1}{a}-1\right)\left(\frac{1}{a}\frac{d\tilde{\psi}}{dx}-\frac{\gamma}{2}a\frac{d\tilde{\varepsilon}}{dx}\right)}{\gamma\sigma^2-\frac{d^2\tilde{\Delta}}{dx^2}+\frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2}+\frac{\gamma}{2}a\frac{d^2\tilde{\varepsilon}}{dx^2}}$$

by a strictly smaller term,

$$\frac{(1-a)\left(\frac{dx}{da}+x\frac{dx}{db}\right)}{(1-a)\frac{dx}{db}+1}=\frac{\left(\frac{1}{a}-1\right)\left(\frac{1}{a}\frac{d\tilde{\psi}}{dx}-\frac{\gamma}{2}a\frac{d\tilde{\varepsilon}}{dx}\right)}{\gamma\sigma^2\left(1+\frac{\lambda_D}{\lambda_M}\right)-\frac{d^2\tilde{\Delta}}{dx^2}+\frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2}+\frac{\gamma}{2}a\frac{d^2\tilde{\varepsilon}}{dx^2}}.$$

Recall that the term in square brackets in (E.8) is strictly positive (by Assumption 1). Therefore in the socially optimal case, there are additional negative terms (or the positive terms are smaller) in the first-order condition with respect to a relative to that in the privately optimal case.

As in the main model in the text, the planner recognizes that incentive provision is weaker than individual fund investors perceive it to be. This is captured by additional negative terms in the first-order conditions for a and b. It is no longer straightforward to establish that the presence of these terms imply that both a and b in the socially optimal case are smaller than those in the privately optimal case. Doing so requires us to impose additional, hard to interpret, assumptions on the cross-derivatives and third derivatives of the functions $\tilde{\Delta}$, $\tilde{\psi}$ and $\tilde{\varepsilon}$. Intuitively, these assumptions are sufficient conditions to guarantee that a and b are complements.

We can still prove the crowded trades result, namely, $p^{**} < p^*$. Define $k = (a, b)$, $W(k,p) = U^F(k,p,x(k,p),e^*(k,x(k,p))) + U^M(k,p,x(k,p),e^*(k,x(k,p)))$. The fund investor's problem is to maximize $W(k,p)$ with respect to k taking p as given. Since we cancel out the distributive effects in the social planner's problem, it is equivalent to maximizing $W(k,p(k))$ with respect to k.

Denote the optimal solutions in the privately and socially optimal cases by k^* and k^{**}, respectively. Notice that

$$W(k^{**}, p(k^{**})) > W(k^{*}, p(k^{*})) > W(k^{**}, p(k^{*}))$$

implying

$$W(k^{**}, p(k^{**})) > W(k^{**}, p(k^{*})). \tag{E.9}$$

Differentiating W with respect to p (and canceling the distributive effects),

$$\frac{dW}{dp} = \frac{\partial U^F}{\partial x}\frac{dx}{dp} = (1-a)\left\{\gamma\sigma^2\left[(2a-1)x(p)-2b\right] + \frac{1}{a}\frac{d\tilde{\psi}}{dx} + \frac{\gamma}{2}(2a-1)\frac{d\tilde{\varepsilon}}{dx}\right\}\frac{dx}{dp} < 0.$$

Differentiating with respect to p one more time,

$$\begin{aligned}\frac{d^2W}{dp^2} &= \frac{dW_s}{dx}\left(\frac{dx}{dp}\right)^2 + W_s\underbrace{\frac{d^2x}{dp^2}}_{=0} \\ &= \left[\gamma\sigma^2(2a-1)x + \frac{1}{a}\frac{d^2\tilde{\psi}}{dx^2} + \frac{\gamma}{2}(2a-1)\frac{d^2\tilde{\varepsilon}}{dx^2}\right]\left(\frac{dx}{dp}\right)^2 > 0\end{aligned}$$

by Assumption 1. Since $dW(k^{**}, p)/dp < 0$ at $p = p^{**}$, this implies that $W(k^{**}, p(k^{**})) < W(k^{**}, p)$ for $p < p(k^{**})$. Given inequality (E.9), it must be the case $p(k^{**}) < p(k^{*})$. It then also follows that $x(k^{**}) < x(k^{*})$. So the crowded trade results from the main text extends to the case with unobservable effort.

E.2 Endogenous Δ

In this appendix we consider the case in which Δ is determined in equilibrium in the market for securities lending. We include a new class of investors who seek to borrow the stock from fund managers so that they could sell them short. These investors therefore incur a borrowing cost of Δ per share, which allows the fund managers to earn revenue of Δ per share. Typical motives for shorting considered in the literature are (i) hedging and (ii) speculation. We choose the first one, so that the resulting model is not too far from our baseline setting. We believe that the insights of this appendix go through in alternative settings, so long as one is not adding market frictions together with additional classes of agents.

We consider a new group of agents, hedgers, H (measure λ_H), endowed with $e\tilde{D}$ units of consumption in period 1.[45] They engage in short selling in period 0 for hedging purposes. Their utility (converted into the mean-variance form) is

$$\max_x \ (x+e)\mu - xp + x\Delta\mathbf{1}_{x\leq 0} - \frac{\gamma}{2}(x+e)^2\sigma^2,$$

where Δ is the borrowing cost and it is incurred only when the hedgers' demand is negative. It is easy to show that the hedgers' portfolio demand is given by

$$x^H = \frac{\mu - p + \Delta}{\gamma\sigma^2} - e. \tag{E.10}$$

We focus on the case when e is large enough so that x^H.

In practice, a fund manager would not be permitted to lend out the entire portfolio and would lend out only a fraction of it. We assume that the number of shares lent out by the manager is ℓx^M, where $\ell \in (0,1]$ is exogenous. The fund's augmented return is now $\ell\Delta x^M$ and the manager's cost is $\ell\psi x^M$. The manager's portfolio is then

$$x^M = \frac{\mu - p + \ell\Delta - \ell\psi/a}{a\gamma\sigma^2} + \frac{b}{a}. \tag{E.11}$$

Substituting (E.10) and (E.11) into the securities-lending market-clearing condition,

$$\ell\lambda_M x^M + \lambda_H x^H = 0, \tag{E.12}$$

[45]Without loss of generality, we assume that the hedgers are endowed with zero shares at time zero.

leads to the following expression for $p - \ell\Delta$:

$$p - \ell\Delta = \mu - \frac{1}{\lambda_H + \ell\lambda_M/a}\left[\gamma\sigma^2\left(\lambda_H e - \ell\lambda_M \frac{b}{a}\right) + \ell\lambda_M \frac{\ell\psi}{a^2} - (1-\ell)\lambda_H \Delta\right]. \tag{E.13}$$

With the new class of agents, the market-clearing condition in the asset market becomes

$$\lambda_M x^M + \lambda_H x^H + \lambda_D x^D = \bar{x},$$

which, using (E.12), can be written as

$$(1-\ell)\lambda_M x^M + \lambda_D x^D = \bar{x}. \tag{E.14}$$

Substituting (3) and (E.11) and solving for $p - \ell\Delta$ yields

$$p - \ell\Delta = \mu - \frac{1}{\lambda_D + (1-\ell)\lambda_M/a}\left[\gamma\sigma^2\left(\bar{x} - (1-\ell)\lambda_M \frac{b}{a}\right) + \lambda_D \ell\Delta - \frac{(1-\ell)\lambda_M}{a}\ell\frac{\psi}{a}\right]. \tag{E.15}$$

Next, we compare the privately and socially optimal contracts. To do this, we consider first-order conditions with respect to b/a and a. The first-order condition for the privately optimal case with respect to b/a and a are

$$\ell\Delta - \ell\psi + \mu - p - \gamma\sigma^2 z = 0 \tag{E.16}$$

and

$$\begin{aligned} 0 &= -(2a-1)\gamma\sigma_\epsilon^2 + \frac{1-a}{a}(\ell\Delta + \mu - p - \gamma\sigma^2 z)\frac{\partial y}{\partial a} \\ &= -(2a-1)\gamma\sigma_\epsilon^2 + (1-a)\ell\frac{\psi^2\sigma^2}{\gamma a^3}, \end{aligned} \tag{E.17}$$

respectively.

Now consider the socially optimal case. Define $U^H = x^H(\Delta + \mu - p) + e\mu - \frac{\gamma}{2}\left(x^H + e\right)^2\sigma^2$.

The social planner's problem is

$$\max_{a,b,c}\ \omega_F U^F + \omega_D U^D + \omega_H U^H$$

subject to (3), (7), (E.10), (E.11), (E.13), and (E.15). Denote

$$y = ax^M - b = \frac{\mu - p + \ell\Delta - \ell\psi/a}{\gamma\sigma^2}.$$

The social planner's first-order condition with respect to b/a is

$$\begin{aligned} 0 = &\left[\omega_F\left(x^F_{-1} - x^M\right) + \omega_D\left(x^D_{-1} - x^D\right) - \omega_H x^H\right]\frac{\partial p}{\partial(b/a)} + \left[\omega_F \ell x^M + \omega_H x^H\right]\frac{\partial\Delta}{\partial(b/a)} \\ &+ \ell\Delta - \ell\psi + \mu - p - \gamma\sigma^2 z + \left(\ell\Delta + \mu - p - \gamma\sigma^2 z\right)\left[\frac{1}{a} - 1\right]\frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial(b/a)}. \end{aligned}$$

As in the main text, we choose the Pareto weights to eliminate the distributive effect. Specifically, if $\omega_F = \lambda_M$, $\omega_D = \lambda_D$, and $\omega_H = \lambda_H$, then the terms in the first line of (E.18) are zero by market clearing. Thus the planner's first-order with respect to b/a becomes

$$\ell\Delta - \ell\psi + \mu - p - \gamma\sigma^2 z + \left(\ell\Delta + \mu - p - \gamma\sigma^2 z\right)\left[\frac{1}{a} - 1\right]\frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial(b/a)} = 0. \tag{E.18}$$

Differentiating (E.13) and (E.15) with respect to b/a, we can solve for $\partial(p-\ell\Delta)/\partial(b/a)$:

$$\frac{\partial(p-\ell\Delta)}{\partial(b/a)} = \Gamma\gamma\sigma^2,$$

where

$$\Gamma = \frac{[\ell^2\lambda_D + (1-\ell)^2\lambda_H]\lambda_M}{\lambda_D\lambda_H + [\ell^2\lambda_D + (1-\ell)^2\lambda_H]\,\lambda_M/a} \in (0,1).$$

and using $\partial y/\partial(p-\ell\Delta) = -1/(\gamma\sigma^2)$, we can rewrite the social planner's first-order condition with respect to b/a as

$$\ell\Delta - \frac{\ell\psi}{1-(1/a-1)\,\Gamma} + \mu - p - \gamma\sigma^2 z = 0. \tag{E.19}$$

The planner's first-order condition with respect to a (after canceling out the distributive effect) is

$$\begin{aligned}0 &= -(2a-1)\gamma\sigma_\epsilon^2 - \left[\ell\Delta - \ell\psi + \mu - p - \gamma\sigma^2 z\right]\frac{y}{a^2}\\ &\quad + \frac{1-a}{a}(\ell\Delta+\mu-p-\gamma\sigma^2 z)\left[\frac{\partial y}{\partial a} + \frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial a}\right],\end{aligned}$$

which, using (E.18), becomes

$$\begin{aligned}0 &= -(2a-1)\gamma\sigma_\epsilon^2\\ &\quad + \frac{1-a}{a}(\ell\Delta+\mu-p-\gamma\sigma^2 z)\left[\frac{\partial y}{\partial a} + \frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial a} + \frac{y}{a^2}\frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial(b/a)}\right].\end{aligned}$$

As in the main model, differentiating (E.14) with respect to b/a and a, we can show that

$$\frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial a} + \frac{y}{a^2}\frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial(b/a)} = \frac{1}{a}\frac{\partial y}{\partial a}\frac{\partial y}{\partial(p-\ell\Delta)}\frac{\partial(p-\ell\Delta)}{\partial(b/a)} = -\frac{\Gamma}{a}\frac{\partial y}{\partial a}.$$

Thus the planner's first-order condition with respect to a is

$$\begin{aligned}0 &= -(2a-1)\gamma\sigma_\epsilon^2 + \frac{1-a}{a}(\ell\Delta+\mu-p-\gamma\sigma^2 z)\left(1-\frac{\Gamma}{a}\right)\frac{\partial y}{\partial a}\\ &= -(2a-1)\gamma\sigma_\epsilon^2 + (1-a)\ell\frac{\psi^2}{\gamma\sigma^2 a^3}\frac{1-\Gamma/a}{1-\Gamma/a+\Gamma}.\end{aligned} \tag{E.20}$$

Comparing (E.19) with (E.16) and (E.20) with (E.17), we can see that the benefit of incentive provision is lower for the planner than for private agents, just as in the main text. The same proofs as in the main model go through for this case and thus our main results continue to hold.

The intuition for why our results go through in this setting is the following. First, all the frictions from the main model are still present. Second, the addition of hedgers and the motive for short selling do not create any additional sources of inefficiency. In particular, adding the hedgers does not complicate the contracting problem. Just as with direct investors, contracts only affect hedgers through the distributive effect. The pecuniary externality occurs because prices (now both p and Δ) enter the manager's incentive constraint. So all the forces are the same as in the main model. The mechanism for alleviating the friction is the same as in the main text, i.e., it involves using skin-in-the-game and benchmarking. The comparison of the privately and socially optimal contracts is also the same.