Informational Hold-up and Performance Persistence in Venture Capital*

Yael V. Hochberg
Northwestern University and NBER

Alexander Ljungqvist
New York University, CEPR, and NBER

Annette Vissing-Jørgensen
Northwestern University, CEPR, and NBER

October 16, 2012

Abstract

Why don't successful venture capitalists eliminate excess demand for their follow-on funds by aggressively raising their performance fees? We propose a theory of learning that leads to informational hold-up in the VC market. Investors in a fund learn whether the VC has skill or was lucky, whereas potential outside investors only observe returns. This gives the VC's current investors hold-up power when the VC raises his next fund: Without their backing, he cannot persuade anyone else to fund him, since outside investors would interpret the lack of backing as a sign that his skill is low. This hold-up power diminishes the VC's ability to increase fees in line with performance. The model provides a rationale for the persistence in after-fee returns documented by Kaplan and Schoar (2005). Empirical evidence from a large sample of U.S. VC funds is consistent with the model. We estimate that up to 68.7% of VC firms lack skill.

*We are grateful to Viral Acharya, Ulf Axelson, April Franco, Rick Green, Arvind Krishnamurthy, Adair Morse, Ludovic Phalippou, Morten Sørensen, Chris Yung, and to seminar participants at the WFA meetings, the NBER Entrepreneurship meetings, the Queen's University Economics of Innovation conference, the Gerzensee meetings, Berkeley, Georgetown, Helsinki, Iowa, LBS, LSE, Miami, Northwestern, NYU, Toronto, Toulouse, Colorado, USC and UC Davis for helpful comments. We would especially like to thank Marco Da Rin and Ludo Phallipou for sharing their LP survey data. Hochberg gratefully acknowledges funding from the Heizer Center for Entrepreneurial Studies and the Zell Center for Risk Research at Kellogg. Ljungqvist gratefully acknowledges generous financial support from the Ewing Marion Kauffman Foundation. We are grateful to VentureOne for providing LP data. Address correspondence to y-hochberg@kellogg.northwestern.edu (Hochberg), al75@nyu.edu (Ljungqvist), or a-vissing@kellogg.northwestern.edu (Vissing-Jørgensen).

The performance of venture capital (VC) funds appears highly persistent across a sequence of funds managed by the same manager (Kaplan and Schoar (2005)). This contrasts with evidence for mutual funds (Malkiel (1995)) and raises an interesting question: Why do successful VCs not raise their fees in line with performance, effectively auctioning off the stakes in their follow-on funds to the highest bidder?

As Berk and Green (2004) show in the context of mutual funds, if investors supply their capital competitively but fund management skill is scarce, investors' expected excess returns must equal zero, realized returns must be unpredictable from public information, and fund managers will earn economic rents reflecting their skill. This fits the mutual fund industry, where returns do not appear persistent, but not the VC industry. Instead, we argue that to explain performance persistence in VC funds, the investor market must become uncompetitive in some way, forcing VCs to share the rents their skills generate with their investors.

A constant level of market power among investors over time is not sufficient to generate persistence. To see why, suppose there is a permanent shortage of investors willing to tie up their capital for the ten-year duration that is common in VC funds. Market power then implies that investors earn positive expected excess returns, by virtue of sharing in the VC's rents, but these expected returns, though positive, will be equal across funds (holding risk constant). Moreover, realized returns must remain unpredictable from public information; otherwise, investors could improve their expected returns by reallocating capital across VCs. Thus, to explain persistence, we need investors' market power to have increased by the time a VC raises his next fund.

We propose a model of learning and informational hold-up in the VC market designed to explain performance persistence. The key unknown is whether a fund manager (the general partner or GP) has skill. To begin with, investors (the limited partners or LPs) do not know the GP's skill, but because skill drives performance, over time LPs have an opportunity to learn. We model GPs as potentially managing a sequence of two funds, each lasting two periods and partially overlapping in time. Thus, a second fund would be raised before the final performance of the GP's first fund is publicly known. Whether a second fund is actually raised depends on what investors have learned about the GP's skill.

The key ingredient of the model is that investing in a fund gives LPs an opportunity to collect 'soft' information about the GP's skill. Other investors in the market, on the other hand, can only observe verifiable 'hard' information, such as realized returns. Access to soft information gives LPs an informational advantage over the market when it comes to distinguishing between skill and luck.[1] Soft information is arguably particularly important in the VC industry: VCs invest in risky, unlisted, and hard-to-value companies which they hold for a number of years before eventually selling them (or, more often, writing them off). Objective returns thus take many years to materialize, unlike in the mutual fund industry where managers invest in traded securities that can easily and objectively be valued, potentially in real time.[2]

It is the asymmetric evolution of information that makes the LP market uncompetitive over time in our model. When a GP raises his first fund, no-one knows his skill. Thus, he faces a large set of potential investors and the LP market is initially perfectly competitive. But over time, as 'incumbent' LPs find out his skill before outside investors do, the LP market becomes less competitive. This asymmetric learning in turn enables incumbent LPs to hold the GP up when he next raises a fund, because other potential investors would interpret failure to reinvest by incumbent LPs as a negative signal about the GP's skill. Specifically, outside investors face a winner's curse—the better-informed incumbent LPs will outbid them in a follow-on fund whenever the GP has skill—and so withdraw from the market for follow-on funds. This gives incumbent LPs bargaining power when negotiating the terms of a follow-on fund with the GP and leads to performance persistence: Net of the fee paid to the GP, high LP returns in a first fund predict high LP returns in a follow-on fund, as the hold-up problem prevents the GP from raising the fee to the

[1] For empirical evidence of the importance of soft information in learning about *corporate* managers' skill, see Cornelli, Kominek, and Ljungqvist (2012).

[2] Lerner, Schoar, and Wongsunwai (2007) note that: "Reinvestment decisions by LPs are particularly important in the private equity industry, where information about the quality of different private equity groups is more difficult to learn and often restricted to existing investors." Lerner and Schoar (2004) argue that LPs typically demand wide-ranging information rights in order to inform their decision whether to reinvest. Chung et al. (2010) use a learning model to calibrate the incentive effects of future fundraising in the VC market.

point where investors just break even.

A natural question to ask is why the GP cannot play off the incumbent LPs in his first fund against each other, such that the LPs compete away the rents when negotiating their investments in his second fund. To allow for such a 'Bertrand equilibrium' outcome, we model each first fund as having two incumbent LPs. As our sequential bargaining model shows, incumbent LPs will be able to hold the GP up, and so enjoy performance persistence, as long as idiosyncratic fund risk is sufficiently high and LPs are sufficiently risk-averse. Intuitively, the combination of risk aversion and idiosyncratic risk implies that LPs effectively behave as if they supply funds at an increasing marginal cost. This prevents them from competing for fund allocations as intensely as they would in a standard Bertrand competition setting (which assumes constant marginal costs).

Both idiosyncratic fund risk and investor risk aversion are plausible features of the VC market. Using data for 1980-2006, we estimate that the dispersion in after-fee returns is 2.5 times greater for VC funds than for mutual funds and 1.4 times greater for VC funds than for hedge funds. To illustrate this point, Figure 1 graphs kernel densities of after-fee IRRs for these three types of funds (as well as for buyout funds, which have a similar risk profile as VC funds). The main reason for the much higher risk of VC funds is that most VC portfolio companies fail. Ljungqvist and Richardson (2003), for example, estimate that as many as three-quarters of portfolio companies are written off in the average fund raised in 1981-1993. From the point of view of an LP, therefore, investing in a VC fund entails a considerable amount of idiosyncratic risk. As we show, this affects the equilibrium outcome if investors are risk averse. Risk aversion, in turn, is a standard assumption in the VC setting; see, for example, Jones and Rhodes-Kropf (2003), Sorensen, Wang, and Yang (2012), and Ang, Papanikolaou, and Westerfield (2012).

Asymmetric learning implies that incumbent LPs and outside investors have different information sets. If learning is indeed asymmetric, proxies for incumbent LPs' 'soft' information should predict not only future performance but also a VC's ability to raise a follow-on fund and the size of the follow-on fund if raised, over and above publicly available 'hard' information. It is this distinction between 'soft' and 'hard' information that allows us to test whether informational hold-up can explain performance persistence in venture capital.

We test the model's predictions using both survey and observational data. The former data come from Da Rin and Phalippou (2011), who conduct a survey of 2,000 LPs in private equity and venture capital funds between 2008 and 2010. Among the questions they ask is the following: "In your experience, does investing in a fund give you priority over other investors when the GP raises subsequent funds?" We tabulate the responses in Table 1. Of the 239 LPs who answered this question, 87.5% indicated receiving priority over outside investors in follow-on funds. Moreover, 72.1% of these LPs agreed with the following statement: "If I didn't re-invest, other investors would be suspicious and would not invest." This directly supports the holdup argument that our model formalizes.

The observational data we use constitutes one of the most comprehensive datasets on U.S. VC funds assembled to date. The data cover 2,257 funds raised by 962 VC firms over the period 1980 to 2002. Unlike Kaplan and Schoar (2005), who have access only to anonymized fund performance data, we know the identity of each fund and each firm in our dataset. This allows us to track each fund and each firm through October 2012. Importantly, we not only have access to the final return a fund earns over its lifetime, but we also know how a fund's performance evolves year-by-year over the course of its life. These 'interim' returns are publicly observable by all potential investors at the time of fundraising and so correspond to the 'hard' information in our model. Final returns, on the other hand, become publicly known only at the end of a fund's life.

How to capture soft information? Most VCs raise their next fund well before the end of their current fund's life. Thus, their current fund's final return is not yet known when they go out fundraising. All the market knows at this point is the current fund's interim return. While the interim return constitutes hard information, by construction it reflects a mixture of objective cash-on-cash returns and subjective unrealized capital gains.[3] Incumbent LPs observe the reported

[3] Blaydon and Horvath (2002) document that absent agreed valuation standards in the VC industry, different VC funds report radically different valuations for the same portfolio companies at a given point in time.

interim return as well, but in our model they also possess soft information, say knowledge of whether the GP's unrealized capital gains are likely to materialize or to evaporate. Soft information allows incumbent LPs to learn the GP's skill and thereby helps them predict the current fund's final return ahead of time. Based on this argument, we treat a current fund's final return (which will be revealed many years later) as a proxy for the soft information that incumbent LPs possess at the time the GP raises his next fund. We are not aware of any previous work with access to both interim and final IRRs.

Our results confirm that VC performance is persistent, consistent with Kaplan and Schoar (2005). Future fund returns are predictable not only based on publicly available 'hard' information but also based on our proxy for 'soft' information, consistent with the predictions of our model. Moreover, 'soft' information also predicts whether a VC can raise a follow-on fund and if so, how much capital he can raise. These results are consistent with the asymmetric learning and so with the economic mechanism at the heart of our model—informational hold-up.

Finally, our data allow us to estimate the prevalence of skill in the VC industry. The model predicts that VCs will go out of business (in the sense of being unable to raise a follow-on fund) once their investors have learned that they lack skill. We find that 661 of the 962 VC firms in our sample (68.7%) go out of business between 1980 and 2012. This suggests that skill is relatively rare in the VC industry. On average, VC firms fail after 14.5 years, having raised 2.7 funds over their lifetime.

Our paper is related to the literature on relationship-banking, which employs similar informational assumptions (e.g. Sharpe (1990), Rajan (1992), von Thadden (2004)), and to the literature on learning more generally in financial markets (see Pastor and Veronesi (2009) for a recent survey). However, unlike in hold-up models in the banking literature, we show that asymmetric learning is in fact efficient in the VC setting. This follows because VC contracts specify both an investment level (fund size) and the division of the fund's surplus between GPs and LPs and, as we show, fund size turns out to be first-best in both first and follow-on funds. Moreover, GPs may even benefit from informational hold-up ex ante, because under certain conditions, first funds can only be raised under asymmetric learning. Such a preference is consistent with the fact that GPs are willing to provide their LPs with considerable amounts of soft information about strategies and performance that cannot credibly be communicated to potential new investors.

In addition, our paper relates to the literatures on VC performance and the relationship between LPs and GPs. Jones and Rhodes-Kropf (2003) provide empirical evidence in support of the hypothesis that VCs need to be compensated for bearing idiosyncratic risk through higher expected returns. Ljungqvist and Richardson (2003) analyze the cash flow, return, and risk characteristics of private equity funds. Cochrane (2005), Korteweg and Sørensen (2010), and Quigley and Woodward (2003) estimate the risk and return of VC investments. Lerner, Schoar, and Wongsunwai (2007) find large heterogeneity in the returns that different classes of investors earn when investing in private equity and suggest that LPs vary in their level of sophistication.

The remainder of the paper is structured as follows. Section I presents our model of learning and informational hold-up. Section II presents the sample and data. Section III presents the empirical analysis, and Section IV discusses and concludes.

I. A Model of Learning About GP Skill

A. Setup

Timeline: At $t = 0$, there is a continuum of GP types of mass one who differ in skill, μ^i. We assume that μ^i is distributed uniformly over the interval $[-\mu, \mu]$, such that $\mu^i = 0$ corresponds to average skill. At this time, no-one knows which GP has skill. We abstract from agency problems by assuming that GPs manage their funds in their LPs' best interest.[4] Each GP raises a "first-time" fund of size I_0. A fund lasts two periods and generates cash flows that will be paid out to investors at $t = 2$, the end of the second period.

[4] For a model of agency problems among fund managers in a learning setting, see Ljungqvist, Richardson, and Wolfenzon (2007).

Between $t = 0$ and $t = 1$, the GP makes a number of investments in portfolio companies whose performance is hit by a random shock, ε^i.

At $t = 1$, the GP publicly releases signal H_1^i, reflecting the fund's interim performance. This can be thought of as the fund's interim IRR at $t = 1$, which in practice would partly consist of unrealized capital gains on illiquid companies that remain in the fund's portfolio at that time.

Between $t = 1$ and $t = 2$, the GP attempts to exit as many portfolio companies as possible, through IPOs or sales. This process is subject to another random shock, v^i.

At $t = 2$, a first-time fund returns a cash flow of $C_2^i = A_2^i \ln\left(1 + I_0^i\right)$ and the GP reports the fund's final IRR, H_2^i.[5] A_2^i captures the effects of the two random shocks and the GP's skill, μ^i:

$$A_2^i = a + f_1(\mu^i, \varepsilon^i) + f_2(\mu^i, v^i) \tag{1}$$

$$\varepsilon^i \sim N\left(0, \frac{\frac{1}{2}\sigma^2\left(I_0^i\right)^2}{\left[\ln\left(1+I_0^i\right)\right]^2}\right),\ v^i \sim N\left(0, \frac{\frac{1}{2}\sigma^2\left(I_0^i\right)^2}{\left[\ln\left(1+I_0^i\right)\right]^2}\right),\ \varepsilon^i \text{ and } v^i \text{ independent.} \tag{2}$$

We parameterize the interim and final performance signals as $H_1^i = f_1(\mu^i, \varepsilon^i) = \mu^i + \varepsilon^i$ and $H_2^i = f_2(\mu^i, v^i) = \mu^i + v^i$. The letter H is used to indicate that these signals constitute *hard* information, i.e., information that can be verified by all parties. The challenge for investors is to disentangle skill from the two stochastic shocks. As we will show, interim and final IRRs can be thought of as noisy signals of the GP's skill.

Depending on information learned between $t = 0$ and $t = 1$, the GP may raise a "follow-on" fund of size I_1 at $t = 1$ which will pay out cash flows two periods later, at $t = 3$. The overlapping timing structure of the model captures real-world practice, by which follow-on funds are typically raised before the first fund has completed its life cycle, i.e., before its final IRR is publicly known.

A follow-on fund, if raised, returns a cash flow of $C_3^i = A_3^i \ln\left(1 + I_1^i\right)$ at $t = 3$. Using subscript

[5] The log function captures decreasing returns to scale. This is similar to Berk and Green's (2004) assumption for mutual funds and consistent with the evidence reported for private equity funds in Lopez de Silanes, Phalippou, and Gottschalg (2010).

"follow-on" to refer to a follow-on fund, we assume that

$$A_3^i = a + H_{2,follow-on}^i + H_{3,follow-on}^i \tag{3}$$

$$H_{2,follow-on}^i = \mu^i + \varepsilon_{follow-on}^i,\ H_{3,follow-on}^i = \mu^i + v_{follow-on}^i \tag{4}$$

$$\varepsilon_{follow-on}^i \sim N\left(0, \frac{\frac{1}{2}\sigma^2\left(I_1^i\right)^2}{\left[\ln\left(1+I_1^i\right)\right]^2}\right),\ v_{follow-on}^i \sim N\left(0, \frac{\frac{1}{2}\sigma^2\left(I_1^i\right)^2}{\left[\ln\left(1+I_1^i\right)\right]^2}\right),$$

where $\varepsilon_{follow-on}^i$ and $v_{follow-on}^i$ are independent of each other and of ε^i and v^i. All shocks are drawn independently across GPs and are independent of the GP's skill, μ^i.[6] For simplicity, all risk is idiosyncratic.

Limited partners: At $t = 0$, there is a large set of identical investors, such that the LP market is perfectly competitive. Each GP chooses two LPs for his first-time fund. Two is sufficient to formally show that the presence of multiple informed investors will not eliminate the informational hold-up that is at the heart of our model, while still preserving mathematical tractability.[7] For simplicity, we assume that GPs do not invest in their own funds.[8] At $t = 1$, we distinguish between *incumbent* LPs, who have invested in a given GP's first-time fund, and *outside* investors, who have not.

Learning about GP skill: At $t = 1$, the GP and the incumbent LPs—but not outside investors—are assumed to have learned the GP's skill, μ^i. Their knowledge of μ^i constitutes soft information which cannot be credibly communicated to third parties as it cannot be objectively verified. Thus, talented GPs cannot credibly convince outside investors of their skill, except to the extent that their skill is noisily reflected in the fund's interim performance signal, H_1^i. Based on observing this signal, outside investors update their beliefs about the GP's skill from the unconditional mean of

[6]The dependence of the variance of A_2^i and A_3^i on fund size is chosen to simplify the analysis by ensuring that the variance of fund returns does not depend on fund size. The normal distribution of cash flows and the uniform distribution of GP types allow us to obtain more closed-form solutions but do not qualitatively drive our results. The more important choice is the functional form of the relation between cash flows and investment, which requires a functional form whereby C_3/I_1 is increasing in GP type μ^i even when I_1 is chosen optimally to reflect GP skill.

[7]While modeling the optimal number of LPs would complicate the model beyond the point of tractability, the intuition for why multiple informed investors will not compete away their hold-up power does not, as we will show, depend on the number of LPs.

[8]In practice, LPs typically contribute 99% of a fund's capital, with the GP providing the remainder.

$E(\mu^i) = 0$ to $E(\mu^i | H_1^i)$.

We refer to this setup as asymmetric learning, in the sense that incumbent LPs learn the GP's type faster than do outside investors.[9] We distinguish this setup from one with symmetric learning in which both incumbent and outside investors learn the GP's type perfectly at $t = 1$.

Preferences and wealth: Both GPs and LPs are risk averse and have CARA preferences over wealth at time $t = 3$, when the cash flow from any follow-on fund is revealed. GPs and LPs have initial wealth of W_0^{GP} and W_0^{LP}, respectively. In addition to investing in the VC industry, LPs can invest at a riskfree rate of r_f, set equal to zero for simplicity. We assume that each LP can invest in one first-time fund and, if desired, in a follow-on fund by the same GP. Cash flows received at $t = 2$ from first-time funds are invested at the riskfree rate from $t = 2$ to $t = 3$.

Payoff functions: Denote the two incumbent LPs in a first-time fund by a and b. We assume that the GP and the LPs divide the fund's cash flow according to the following contract. At $t = 2$, the GP is paid a dollar amount of[10]

$$X_0^{GP} = M_{0,a} + M_{0,b} \equiv 2M_0 \tag{5}$$

while the two LPs each receive cash flows net of fees equal to

$$X_0^{LP} = C_2^i/2 - M_0. \tag{6}$$

As we will see shortly, follow-on funds have either one or two LPs. If both incumbent LPs a and b invest in the follow-on fund, the GP receives a fee of $M_{1,split,a}$ from LP_a and $M_{1,split,b}$ from LP_b. If only one of them invests, the GP receives either $M_{1,sole,a}$ or $M_{1,sole,b}$, depending on who invests. The values of the fee, the fund size, and the number of LPs who invest in the follow-on

[9]Having incumbent LPs learn the GP's skill perfectly at $t = 1$ is stronger than necessary. All that is required for our results to go through is that incumbent LPs receive a more precise signal at $t = 1$ than do outside investors.

[10]Fees in VC contracts are usually expressed in percentage terms. For tractability, we model fees in dollars. Once the optimal fund size has been derived, one can easily calculate the implied percentage fee.

fund will be the focus of the solution of the model.

We abstract from performance fees. In practice, GPs are paid both a fixed management fee (as modeled here) and a performance fee (in the form of the carried interest or "carry"). The latter is intended to provide the GP with incentives to exert effort. As our model abstracts from effort provision, there is no need to include a performance component in the contract.

B. Fund Size and Fee in Follow-On Funds

Under asymmetric learning, the LP market is perfectly competitive at $t = 0$ but not at $t = 1$. Because outside investors have not learned the GP's type fully when the GP attempts to raise a follow-on fund, incumbent LPs have an informational advantage. This will allow incumbent LPs to extract part of the follow-on fund's value from the GP.

While it is intuitive that the informational advantage of incumbent LPs should improve the terms they obtain, it is useful to model the bargaining game explicitly, for two reasons. First, it will allow us to show that the presence of *multiple* incumbent LPs does *not* eliminate the informational hold-up that is our central mechanism for generating performance persistence. This will be the case as long as LPs are sufficiently risk averse and idiosyncratic fund risk is sufficiently high. Second, explicitly modeling the bargaining allows us to be clear about the role played by outside investors.

Bargaining: We extend Rubinstein (1982) bargaining to a setting with three parties and risk aversion. Starting at $t = 1$, the GP and incumbent LPs a and b bargain sequentially as follows:

(i) The GP makes an offer to LP_a and LP_b for each to invest $I^{GP}_{1,split}/2$ and for each to pay a fee of $M^{GP}_{1,split}$, for a total fund size of $I^{GP}_{1,split}$ and a total fee of $2M^{GP}_{1,split}$. We denote this as a *split* offer. As an alternative to the split offer, the GP also offers to have a single LP invest $I^{GP}_{1,sole}$ with a total fee of $M^{GP}_{1,sole}$. We denote this as a *sole* offer. The GP's overall offer is hence $\left[(I^{GP}_{1,split}/2, M^{GP}_{1,split}), (I^{GP}_{1,sole}, M^{GP}_{1,sole})\right]$.

(ii) If the GP's offer is rejected, LP_a and LP_b simultaneously counter the GP's offer. LP_a offers to provide either half of the capital needed (a split offer) and pay a fee of $M^{LP_a}_{1,split}$, or to provide all of the capital needed (a sole offer) and pay a fee of $M^{LP_a}_{1,sole}$. This offer is denoted $\left[(I^{LP_a}_{1,split}/2, M^{LP_a}_{1,split}), (I^{LP_a}_{1,sole}, M^{LP_a}_{1,sole})\right]$. Similarly, LP$_b$'s offer is $\left[(I^{LP_b}_{1,split}/2, M^{LP_b}_{1,split}), (I^{LP_b}_{1,sole}, M^{LP_b}_{1,sole})\right]$.[11] The GP can either accept both LPs' split offers, or accept one of the sole offers, or reject both offers.

(iii) If the LPs' offers are rejected, the GP makes another offer; and so on.

We assume that delay in reaching an agreement is costly. Following convention (e.g., Binmore, Rubinstein, and Wolinsky (1986)), we capture this by assuming that between each round of offers, there is an exogenous probability p that the bargaining process will terminate without an agreement.[12]

If no agreement is reached, each party receives its outside option. For the incumbent LPs, this equals the riskfree return, r_f. The GP's outside option depends on what outside investors are willing to offer if no agreement has been reached. We assume that outside investors cannot observe (or at least cannot verify) the bids made prior to bargaining breaking down. Therefore, they do not know whether bargaining has broken down for exogenous reasons or because one of the parties has simply refused to bargain any further. We furthermore assume that an incumbent LP can always counter any offer an outside investor makes.

The GP's outside option is then zero, because outside investors face a winner's curse: Any outside offer would only be accepted if it *reduced* outside investors' expected utility. Why? If the outside offer resulted in a gain in expected utility for the investor who made it, it would immediately be countered by an incumbent LP, who could increase the fee offered to the GP slightly and still enjoy an increase in his own expected utility. As a result, an outside offer would only be successful if the GP's type was sufficiently low so that the incumbents chose not to counter the offer. Outside

[11]Restricting incumbent LP offers to supply either half or all of the capital for the follow-on fund simplifies the analysis while allowing the LPs to compete against each other. Given the parallels between our setup and the procurement setup of Anton and Yao (1989), our results should be robust to allowing for splits other than a half. This is because each LP effectively can veto any split other than the most collusive one by offering the GP a very unattractive fee for providing his share of the funds. See Anton and Yao (1989, p. 539) for an example that shows that results on supplier collusion do not hinge on restricting offers to be for either half or all of the amount.

[12]In the VC setting, this could capture the possibility that the GP's network contacts become stale while he is fundraising, or that another GP makes deals with the relevant entrepreneurs.

investors will therefore rationally withdraw from the market.

The fund size that maximizes joint surplus: To solve for the Nash equilibrium strategies, we first derive the optimal follow-on fund sizes in the split and sole cases, i.e., the fund sizes that maximize the joint surplus of the GP and LPs in each case. These will depend on the GP's skill, μ^i, which we denote using superscript i: $I^i_{1,split}$ and $I^i_{1,sole}$.

Split case: $I^i_{1,split}$ solves

$$\max_{I_1} E\left(U^{GP}|\mu^i\right) + 2E\left(U^{LP}|\mu^i\right) \tag{7}$$

where

$$E\left(U^{GP}|\mu^i\right) = 1 - e^{-\gamma W_3^{GP}} = 1 - e^{-\gamma\left[W_0^{GP}+2M_0+2M_1\right]} \tag{8}$$

$$E\left(U^{LP}|\mu^i\right) = 1 - E\left(e^{-\gamma W_3^{LP}}|\mu^i\right) \tag{9}$$

As we show in Appendix A, this yields an optimal fund size of

$$I^i_{1,split} = \frac{E\left(A_3|\mu^i\right)}{1+\gamma\frac{1}{2}\sigma^2 I^i_{1,split}} - 1 \tag{10}$$

$$= \frac{-\left(1+\gamma\frac{1}{2}\sigma^2\right) + \sqrt{\left(1+\gamma\frac{1}{2}\sigma^2\right)^2 - 2\gamma\sigma^2\left[1 - E\left(A_3|\mu^i\right)\right]}}{\gamma\sigma^2}. \tag{11}$$

Sole case: If only one LP invests in the follow-on fund, maximizing the joint surplus implies

$$\max_{I_1} E\left(U^{GP}|\mu^i\right) + E\left(U^{LP}|\mu^i\right). \tag{12}$$

As we show in Appendix A, this yields an optimal fund size of

$$I^i_{1,sole} = \frac{E\left(A_3|\mu^i\right)}{1+\gamma\sigma^2 I^i_{1,sole}} - 1 \tag{13}$$

$$= \frac{-\left(1+\gamma\sigma^2\right) + \sqrt{\left(1+\gamma\sigma^2\right)^2 - 4\gamma\sigma^2\left[1 - E\left(A_3|\mu^i\right)\right]}}{2\gamma\sigma^2}. \tag{14}$$

Discussion: Regardless of whether one or two LPs invest in the follow-on fund, the optimal fund size does not depend on the fee, M_1. Instead, M_1 simply determines how the surplus is shared. The optimal fund size does, however, depend on the number of LPs in the follow-on fund. The term $\gamma\frac{1}{2}\sigma^2 I^i_{1,split}$ is the risk adjustment to the cost of capital in the split case. It is only half as large as the risk-adjustment to the cost of capital in the case of a single LP, $\gamma\sigma^2 I^i_{1,sole}$. This implies that $I^i_{1,split} > I^i_{1,sole}$. Finally, regardless of the number of LPs who invest, the optimal fund size increases in GP skill (as reflected in $E\left(A_3|\mu^i\right)$) and decreases in risk aversion γ and risk σ^2.

Both $I^i_{1,split}$ and $I^i_{1,sole}$ equal zero for $E\left(A_3|\mu^i\right) = 1$. Since $E\left(A_3|\mu^i\right) = a + 2\mu^i$, this implies that the cutoff GP type for a follow-on fund generating no joint surplus is given by $a + 2\mu^i = 1 \Longleftrightarrow \mu^i = \frac{1-a}{2}$. We denote this value of μ^i by μ^*.

Nash equilibrium strategies, fund size, and fee: The following proposition states the equilibrium outcome of the bargaining game for sufficiently high risk aversion and fund risk.

Proposition 1: Define

$$\begin{aligned} M_1^*\left(\mu^i\right) &= -\ln\left(x\left(\mu^i\right)\right)/\gamma && (15)\\ x\left(\mu^i\right) &= \text{the real root of the cubic equation} \\ & \quad 2e^{b(\mu_i)}x(\mu_i)^3 - x(\mu_i)^2 = 0 \text{ (as derived in Appendix A)} && (16)\\ b\left(\mu^i\right) &= \gamma\frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right) - I^i_{1,split}\right] - \frac{1}{8}\gamma^2\sigma^2\left(I^i_{1,split}\right)^2. && (17) \end{aligned}$$

Then, provided that

$$\begin{aligned} & \frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right) - I^i_{1,split}\right] - \frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2 - M_1^*\left(\mu^i\right) && (18)\\ > \; & \left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,sole}\right) - I^i_{1,sole}\right] - \frac{1}{2}\gamma\sigma^2\left(I^i_{1,sole}\right)^2 - 2M_1^*\left(\mu^i\right) \end{aligned}$$

as $p \longrightarrow 0$, the following is a subgame perfect equilibrium:

(a) All offers involve fund sizes that maximize the joint surplus given the number of LPs investing:

$I^{GP}_{1,split}$, $I^{LP_a}_{1,split}$, and $I^{LP_b}_{1,split}$ all equal $I^i_{1,split}$, and $I^{GP}_{1,sole}$, $I^{LP_a}_{1,sole}$, and $I^{LP_b}_{1,sole}$ all equal $I^i_{1,sole}$ (and $I^i_{1,split}$ and $I^i_{1,sole}$ are zero for $\mu^i < \mu^*$).

(b) LPs prefer $\left[I^i_{1,split}/2, M^*_1\left(\mu^i\right)\right]$ to $\left[I^i_{1,sole}, 2M^*_1\left(\mu^i\right)\right]$.

(c) Denote by $M^{LP,*}_{1,split}$ and $M^{GP,*}_{1,split}$ the fees (paid by each LP) such that (i) LP_a and LP_b are indifferent between accepting the GP's split offer and having their own split offers accepted in the next round and (ii) the GP is indifferent between accepting the LPs' split offers and having his own split offer accepted in the next round. As $p \longrightarrow 0$, $M^{LP,*}_{1,split}$ and $M^{GP,*}_{1,split}$ both converge to $M^*_1\left(\mu^i\right)$.

(d) The GP's strategy is to always offer $\left[(I^i_{1,split}/2, M^{GP,*}_{1,split}), (I^i_{1,sole}, 2M^{GP,*}_{1,split})\right]$ and always reject offers that imply total fees below $2M^{LP,*}_{1,split}$. LP_a and LP_b follow identical strategies. Each of them always offers $\left[(I^i_{1,split}/2, M^{LP,*}_{1,split}), (I^i_{1,sole}, 2M^{LP,*}_{1,split})\right]$ whenever it is the LPs' turn to make an offer and always rejects offers that imply total fees above $2M^{GP,*}_{1,split}$.

Given (b) and (d), the equilibrium outcome of the bargaining game is immediate agreement with both LPs accepting the GP's first split offer. The total fee is thus $2M^*_1\left(\mu^i\right)$, the fund size is $I^i_{1,split}$, and each LP invests $I^i_{1,split}/2$ and pays fees of $M^*_1\left(\mu^i\right)$.

Proof of Proposition 1: See Appendix A.

Corollary 1: For given skill μ^i and thus $E\left(A_3|\mu^i\right)$, condition (18) in Proposition 1 is satisfied for γ and σ^2 sufficiently high. Specifically, for a given μ^i and thus $E\left(A_3|\mu^i\right)$, there exists a function $\gamma^*\left(\sigma^2, E\left(A_3|\mu^i\right)\right)$ that is monotonically decreasing in σ^2 such that deviating is not optimal for $\gamma > \gamma^*\left(\sigma^2, E\left(A_3|\mu^i\right)\right)$. Furthermore, $\gamma^*\left(\sigma^2, E\left(A_3|\mu^i\right)\right)$ is increasing in $E\left(A_3|\mu^i\right)$, which implies that higher values of γ and σ^2 are needed for the condition to be satisfied for GPs with greater skill.

Corollary 1 is represented graphically in Figure 2, which depicts the values of γ and σ^2 for which the condition in Proposition 1 holds, for various values of $E\left(A_3|\mu^i\right)$.

Discussion: Condition (18) in Proposition 1 is intuitive. It compares an LP's risk-adjusted cash flows after fees in the split case to the case in which the LP instead becomes the sole investor. Being the sole investor involves paying twice the fee (plus an epsilon amount to get the GP to prefer having a sole LP) and bearing more risk, but allows the LP to provide all the capital for the fund rather than splitting it with another LP. The fund size will be smaller with a sole investor, as the idiosyncratic risk is then borne by a single LP, and condition (18) reflects this. Corollary 1 clarifies when the condition will hold. Not surprisingly, this will be the case when risk aversion or idiosyncratic fund risk is high, in which case earning all of the fund's cash flow, rather than half of it, fails to compensate for bearing the additional risk and paying all of the fees.[13]

Corollary 2: Proposition 1 implies that LPs earn positive risk-adjusted cash flows in follow-on funds, even after fees, and that these risk-adjusted cash flows increase in the GP's skill, μ^i.

We prove Corollary 2 in Appendix A.

As we will show shortly, the risk-adjusted return after fees (i.e., the cash flow after fees divided by the amount invested) also increases in μ^i. This, in turn, is what generates persistence.

C. Fund Size and Fee in First-Time Funds

As no learning has taken place yet, the LP market is perfectly competitive at $t = 0$. Thus, LPs have no bargaining power and all GPs offer LPs contracts that give the GP the maximum expected utility, subject to each LP achieving an expected utility (across investing in both first and follow-on fund, if raised) that equals the LP's outside option. We refer to this as the LPs' participation constraint. We proceed under the assumption that condition (18) in Proposition 1 holds.

With two LPs investing in both funds raised by a given GP, we have

$$W_3^{LP} = W_0^{LP} + \frac{1}{2}\left(A_2 \ln\left(1 + I_0\right) - I_0\right) + \frac{1}{2}\left[A_3 \ln\left(1 + I_{1,split}^i\right) - I_{1,split}^i\right] - M_0 - M_1^*\left(\mu^i\right) \qquad (19)$$

and

$$W_3^{GP} = W_0^{GP} + 2M_0 + 2M_1^*\left(\mu^i\right). \qquad (20)$$

[13]While Appendix A proves that the above is *a* subgame perfect equilibrium, we cannot prove uniqueness. If one restricted GP and LP strategies to split offers, then the equilibrium in Proposition 1 would be unique, following the argument given in the proof of Result 1 in Binmore, Osborne, and Rubinstein (1992). However, the possibility that the parties can make sole offers complicates the setting so that we are unable to prove uniqueness.

We first determine the LPs' participation constraint. We then solve for the fund size that maximizes the GP's expected utility subject to this constraint. Not surprisingly, the fund size that results will be the one that maximizes joint GP and LP surplus, as was the case for follow-on funds.

LPs' participation constraint: As of $t = 0$, the GP's skill is unknown, and so I_0 will not depend on μ^i. When calculating the LPs' expected utility, however, expectations must be taken both with respect to μ^i and to the shocks A_2 and A_3. Furthermore, follow-on funds are only raised for GP skill $\mu^i > \mu^*$, and thus expectations need to be taken accordingly.

Denote $\frac{1}{2}\left(A_2 \ln\left(1+I_0\right) - I_0\right)$ by Z_2^{LP} and $\frac{1}{2}\left(A_3 \ln\left(1+I_{1,split}^i\right) - I_{1,split}^i\right)$ by $Z_3^{LP}\left(\mu^i\right)$. $Z_3^{LP}\left(\mu^i\right)$ is zero for $\mu^i \leq \mu^*$. Then

$$\begin{aligned} EU^{LP} &= E\left(1-e^{-\gamma W_3^{LP}}\right) = E_{\mu^i}\left(E\left(1-e^{-\gamma W_3^{LP}}|\mu^i\right)\right) \qquad (21) \\ &= 1-e^{-\gamma W_0^{LP}} E_{\mu^i}\left(e^{-\gamma\left[E\left(Z_2^{LP}|\mu^i\right)-M_0\right]+\frac{1}{2}\gamma^2 V\left(Z_2^{LP}|\mu^i\right)} e^{-\gamma\left[E\left(Z_3^{LP}|\mu^i\right)-M_1^*\left(\mu^i\right)\right]+\frac{1}{2}\gamma^2 V\left(Z_3^{LP}|\mu^i\right)}\right) \\ &= 1-e^{-\gamma W_0^{LP}} E_{\mu^i}\left(e^{-\gamma\left[E\left(Z_2^{LP}|\mu^i\right)-M_0\right]+\frac{1}{2}\gamma^2 V\left(Z_2^{LP}|\mu^i\right)} e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}\right), \end{aligned}$$

exploiting that $cov\left(Z_2^{LP}, Z_3^{LP}|\mu^i\right) = 0$. The LPs' participation constraint is that $EU^{LP} = 1 - e^{-\gamma W_0^{LP}}$, i.e., that

$$\begin{aligned} 1 &= E_{\mu^i}\left(e^{-\gamma\left[E\left(Z_2^{LP}|\mu^i\right)-M_0\right]+\frac{1}{2}\gamma^2 V\left(Z_0^{LP}|\mu^i\right)} e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}\right) \Longleftrightarrow \\ M_0\left(I_0\right) &= -\frac{1}{\gamma}\ln E_{\mu^i}\left(e^{-\gamma E\left(Z_2^{LP}|\mu^i\right)+\frac{1}{2}\gamma^2 V\left(Z_2^{LP}|\mu^i\right)} e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}\right) \qquad (22) \end{aligned}$$

where

$$\begin{aligned} E\left(Z_2^{LP}|\mu^i\right) &= \frac{1}{2}\left(E\left(A_2|\mu^i\right)\ln\left(1+I_0\right) - I_0\right) \text{ for any } \mu^i \qquad (23) \\ V\left(Z_2^{LP}|\mu^i\right) &= \frac{1}{4}\sigma^2 I_0^2 \text{ for any } \mu^i \qquad (24) \end{aligned}$$

and

$$b\left(\mu^i\right)=\gamma\frac{1}{2}\left(E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)-\frac{1}{4}\sigma^2\left(I^i_{1,split}\right)^2 \text{ for } \mu^i>\mu^* \text{ and zero otherwise.}$$

The LPs' participation constraint simply says that, to the extent that LPs (due to their informational hold-up power) earn a positive risk-adjusted cash flow after fees in follow-on funds $\left(b\left(\mu^i\right)/\gamma-M_1^*\left(\mu^i\right)>0\right)$, first-time funds must contribute negatively to expected utility.

First-fund size: The GP picks I_0 to maximize his expected utility subject to the LPs' participation constraint:

$$\max_{I_0} E_{\mu^i}\left(1-e^{-\gamma W_3^{GP}}\right) \text{ s.t. } M_0=M_0\left(I_0\right). \tag{25}$$

As $E_{\mu^i}\left(1-e^{-\gamma W_3^{GP}}\right)=1-e^{-\gamma W_0^{GP}}e^{-\gamma 2M_0}E_{\mu^i}\left(e^{-\gamma 2M_1^*\left(\mu^i\right)}\right)$, and since $M_1^*\left(\mu^i\right)$ from Proposition 1 does not depend on what happens in the first fund, this implies simply choosing the value of I_0 that maximizes $M_0\left(I_0\right)$, or equivalently

$$\min_{I_0} E_{\mu^i}\left(e^{-\gamma E\left(Z_2^{LP}|\mu^i\right)+\frac{1}{2}\gamma^2 V\left(Z_2^{LP}|\mu^i\right)}e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}\right). \tag{26}$$

This is equivalent to choosing I_0 to maximize the joint surplus without constraints, since

$$\begin{aligned}
& EU^{GP}+2EU^{LP} \\
= \; & 1-e^{-\gamma W_0^{GP}}e^{-\gamma 2M_0}E_{\mu^i}\left(e^{-\gamma 2M_1^*\left(\mu^i\right)}\right)+ \\
& 2\left[1-e^{-\gamma W_0^{LP}}e^{\gamma M_0}E_{\mu^i}\left(e^{-\gamma\left[E\left(Z_2^{LP}|\mu^i\right)-M_0\right]+\frac{1}{2}\gamma^2 V\left(Z_2^{LP}|\mu^i\right)}e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}\right)\right]
\end{aligned} \tag{27}$$

of which only the last expectation depends on I_0. This term can be rewritten as

$$
\begin{aligned}
& E_{\mu^i}\left(e^{-\gamma E\left(Z_2^{LP}|\mu^i\right)+\frac{1}{2}\gamma^2 V\left(Z_2^{LP}|\mu^i\right)}e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}\right) \\
= \; & \frac{1}{2\mu}\int_{\mu^*}^{\mu} e^{-\gamma\frac{1}{2}\left(E\left(A_2|\mu^i\right)\ln(1+I_0)-I_0\right)+\frac{1}{8}\gamma^2\sigma^2 I_0^2}\left[e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}-1\right]d\mu^i \\
& +\frac{1}{2\mu}\int_{-\mu}^{\mu} e^{-\gamma\frac{1}{2}\left(E\left(A_2|\mu^i\right)\ln(1+I_0)-I_0\right)+\frac{1}{8}\gamma^2\sigma^2 I_0^2}d\mu^i
\end{aligned} \tag{28}
$$

Thus, the first-order condition for the optimal first-fund size, I_0, is:

$$
\begin{aligned}
& \int_{-\mu}^{\mu} e^{-\gamma\frac{1}{2}\left(E\left(A_2|\mu^i\right)\ln(1+I_0)-I_0\right)+\frac{1}{8}\gamma^2\sigma^2 I_0^2}\left\{\left(\frac{E\left(A_2|\mu^i\right)}{(1+I_0)}-1\right)-\gamma\frac{1}{2}\sigma^2 I_0\right\}d\mu^i+ \\
& \int_{\mu^*}^{\mu} e^{-\gamma\frac{1}{2}\left(E\left(A_2|\mu^i\right)\ln(1+I_0)-I_0\right)+\frac{1}{8}\gamma^2\sigma^2 I_0^2}\left\{\left(\frac{E\left(A_2|\mu^i\right)}{(1+I_0)}-1\right)-\gamma\frac{1}{2}\sigma^2 I_0\right\}\left[e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}-1\right]d\mu^i \\
= \; & 0
\end{aligned} \tag{29}
$$

The first integral in the first-order condition captures the optimal first-fund size, considering the fund in isolation. The second term is needed because the presence of a follow-on fund (for $\mu^i > \mu^*$) affects the optimal first-fund size. Using the expression for M_1^* derived in Appendix A, we can show that the term $e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}-1$ in the second integral is always negative for $\mu^i > \mu^*$; goes to 0 for $b\left(\mu^i\right)\to 0$; and goes to -1 for $b\left(\mu^i\right)\to\infty$. As a result, the optimal value of I_0 (denoted I_0^*) is smaller than the value (denote it I_0^x) that would result if I_0 was chosen without consideration of the follow-on fund.[14] Intuitively, the marginal value of increasing fund size I_0 is reduced by the fact that risk-adjusted cash flows are unconditionally (i.e., absent knowledge of μ^i at $t=0$) positively correlated across a GP's two funds, both of whose cash flows increase in μ^i.

First fund fee: While the optimal size of a first fund, I_0^*, cannot be derived in closed form, its

[14] I_0^x solves $\int_{-\mu}^{\mu} e^{-\gamma\frac{1}{2}E\left(A_2|\mu^i\right)\ln(1+I_0)}\left\{\frac{E\left(A_2|\mu^i\right)}{(1+I_0)}-1-\gamma\frac{1}{2}\sigma^2 I_0\right\}d\mu^i=0$. Since $e^{-b\left(\mu^i\right)+\gamma M_1\left(\mu^*\right)}-1<0$, the derivative in the first-order condition is negative at I_0^x, so $I_0^* < I_0^x$.

fee, for any given I_0, can be determined directly from the LPs' participation constraint:

$$
\begin{aligned}
M_0\left(I_0\right) &= -\frac{1}{\gamma}\ln E_{\mu^i}\left(e^{-\gamma E\left(Z_2^{LP}|\mu^i\right)+\frac{1}{2}\gamma^2 V\left(Z_2^{LP}|\mu^i\right)}e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}\right) \\
&= -\frac{1}{\gamma}\ln\left(\begin{array}{c}\frac{1}{2\mu}\int_{\mu^*}^{\mu} e^{-\gamma\frac{1}{2}\left(E\left(A_2|\mu^i\right)\ln(1+I_0)-I_0\right)+\frac{1}{8}\gamma^2\sigma^2 I_0^2}\left[e^{-b\left(\mu^i\right)+\gamma M_1^*\left(\mu^i\right)}-1\right]d\mu^i \\ +\frac{1}{2\mu}\int_{-\mu}^{\mu} e^{-\gamma\frac{1}{2}\left(E\left(A_2|\mu^i\right)\ln(1+I_0)-I_0\right)+\frac{1}{8}\gamma^2\sigma^2 I_0^2}d\mu^i\end{array}\right) \qquad (30)
\end{aligned}
$$

D. Performance Persistence

We can now show that our model implies persistence in LP returns after fees. We then derive additional empirical predictions that should hold if our hold-up model is the correct mechanism underlying these return patterns. We focus on the case where risk aversion and idiosyncratic risk are sufficiently high such that Proposition 1 holds.

Definitions: Let $r^i_{first,final}$ denote the realized after-fee return to LPs in GP i's first fund at $t=2$,

$$r^i_{first,final} = \frac{C_2^i - 2M_0}{I_0} - 1,$$

and let $r^i_{follow-on,final}$ denote the realized after-fee return LPs earn in GP i's follow-on fund at $t=2$,

$$r^i_{follow-on,final} = \frac{C_3^i - 2M_1^*\left(\mu^i\right)}{I_1^i} - 1.$$

The interim return on first funds, $r^i_{first,interim}$, is the after-fee return LPs expect to earn in a first fund of a given GP i, based solely on hard information observable at $t=1$. It is given by

$$r^i_{first,interim} = \frac{E(C_2^i|H_1^i) - 2M_0}{I_0} - 1,$$

where we omit a superscript i on I_0 since it is identical for all GPs.

Implication 1: Persistence in after-fee returns to LPs

(a) In the cross-section of GPs with follow-on funds, a high interim first-fund return predicts a high final return to the LPs in the GP's follow-on fund: $E\left(r^i_{follow-on,final}|r^i_{first,interim},\mu^i > \mu^*\right)$

increases in $r^i_{first,interim}$.

(b) This is true even after adjusting for idiosyncratic risk: $E\left(r^i_{follow-on,final}|r^i_{first,interim}, \mu^i > \mu^*\right) - E\left(\gamma \frac{1}{4}\sigma^2 I^i_{1,split}|r^i_{first,interim}, \mu^i > \mu^*\right)$ increases in $r^i_{first,interim}$.[15]

Implication 1 is, of course, what the model is designed to capture. The proof is presented in Appendix A. One might think that outside investors could simply invest in all follow-on funds raised by GPs who have high $r^i_{first,interim}$, thus expecting to earn high risk-adjusted returns on those follow-on funds. Our model shows why this is not feasible. The winner's curse problem described earlier implies that outside investors would only be able to invest with those GPs for whom their offers implied a reduction in expected utility to investors. This implies that the 'return-chasing' behavior emphasized by Berk and Green (2004) as the mechanism eliminating performance persistence in mutual funds breaks down in the VC setting when there is asymmetric learning.

E. Additional Empirical Implications

In addition to performance persistence, the model yields further empirical implications concerning the probability that a follow-on fund is raised, what its size will be, as well as its expected return.

Implications 2a and 3a below concern the impact of learning on fundraising and fund size and hold whether learning is symmetric or asymmetric. (We derive the outcome of the model for the symmetric learning case in the proof of Implication 2a in Appendix A.) Implications 2b, 3b and 4, on the other hand, hold only if learning is asymmetric and so can be used to test the model against a generic learning story.

In each of the following implications, $r^i_{first,interim}$ directly captures the hard information available to outside investors at the time of follow-on fundraising, while $r^i_{first,final}$ serves as a proxy for incumbent LPs' soft information (i.e., their knowledge of μ^i).

Implication 2: GP fundraising

[15] The risk-adjustment is defined as the reduction in expected return such that each LP would be indifferent between earning the actual $r^i_{follow-on,final}$ and earning a riskless return equal to $E\left(r^i_{follow-on,final}|r^i_{first,interim}, \mu^i > \mu^*\right) - E\left(\gamma \frac{1}{4}\sigma^2 I^i_{1,split}|r^i_{first,interim}, \mu^i > \mu^*\right)$.

(a) Whether or not learning is asymmetric, the probability that a GP raises a follow-on fund increases in the interim return to LPs on the GP's first fund:

(i) Under asymmetric learning, $P\left(\mu^i > \mu^* | r^i_{first,interim}\right)$ increases in $r^i_{first,interim}$.

(ii) Under symmetric learning, $P\left(H^i_1 > \frac{1-a}{2} | r^i_{first,interim}\right)$ increases in $r^i_{first,interim}$.

(b) If learning is asymmetric, soft information about GP skill helps predict if a follow-on fund is raised, over and above the hard information available to outside investors: $P\left(\mu^i > \mu^* | r^i_{first,interim}, r^i_{first,final}\right)$ increases in $r^i_{first,final}$.

Implication 3: Follow-on fund size

(a) Whether or not learning is asymmetric, in the cross-section of GPs with follow-on funds, a high interim return to the LP in the first fund predicts a larger follow-on fund:

(i) Under asymmetric learning, $E\left(I^i_1 | r^i_{first,interim}, \mu^i > \mu^*\right)$ increases in $r^i_{first,interim}$.

(ii) Under symmetric learning, $E\left(I^i_1 | r^i_{first,interim}, H^i_1 > \frac{1-a}{2}\right)$ increases in $r^i_{first,interim}$.

(b) If learning is asymmetric, soft information about GP skill predicts follow-on fund size, over and above the hard information available to outside investors: $E\left(I^i_1 | r^i_{first,interim}, r^i_{first,final}, \mu^i > \mu^*\right)$ increases in $r^i_{first,final}$.

Implication 4: Follow-on fund returns: If learning is asymmetric, soft information about GP skill helps predict LP returns in the GP's follow-on fund, over and above the hard information available to outside investors: $E\left(r^i_{follow-on,final} | r^i_{first,interim}, r^i_{first,final}, \mu^i > \mu^*\right)$ increases in $r^i_{first,final}$.

We prove these implications formally in Appendix A. The intuition for these results is straightforward. The reason that Implications 2a and 3a hold regardless of whether learning is symmetric or asymmetric is that they are independent of how the GP and LPs split the surplus of follow-on funds. They simply follow from the fact that $r^i_{first,interim}$ is informative about the GP's skill, μ^i, and μ^i in turn determines both whether a follow-on fund is raised and its size if raised.

It is Implications 2b, 3b, and 4 which potentially allow us to discriminate between symmetric and asymmetric learning and so to test our model. The intuition for Implication 2b is as follows. Incumbent LPs learn the GP i's type μ^i and the GP can only raise a follow-on fund if they learn that $\mu^i > \mu^*$. This implies that any variable that contains information (to the econometrician) about what incumbents have learned about μ^i helps predict whether a follow-on fund is raised. Specifically, note that the interim return $r^i_{first,interim}$ is an increasing function of the hard information released at $t = 1$, H^i_1 :

$$1 + r^i_{first,interim} = \frac{\frac{1}{2}E(C^i_2|H^i_1) - M_0}{\frac{1}{2}I_0} = \frac{\frac{1}{2}\left(a + 2H^i_1\right)\ln(1 + I_0) - M_0}{\frac{1}{2}I_0}$$

The final return $r^i_{first,final}$ is an increasing function of both the hard information released at $t = 1$, H^i_1, and the additional signal H^i_2 that becomes public information at $t = 2$. Because both signals are functions of skill (i.e., $H^i_1 = \mu^i + \varepsilon^i$ and $H^i_2 = \mu^i + v^i$), we have that

$$1 + r^i_{first,final} = \frac{\frac{1}{2}C^i_2 - M_0}{\frac{1}{2}I_0} = \frac{\frac{1}{2}\left(a + H^i_1 + H^i_2\right)\ln(1 + I_0) - M_0}{\frac{1}{2}I_0}.$$

This implies that $r^i_{first,interim}$ fully reveals H^i_1, and given H^i_1, $r^i_{first,final}$ fully reveals H^i_2. In short, both H^i_1 and H^i_2 (and thus both $r^i_{first,interim}$ and $r^i_{first,final}$) are noisy signals (to the econometrician) about GP type μ^i and thus both are informative for predicting whether incumbents did in fact learn that $\mu^i > \mu^*$.

The intuition for Implications 3b and 4 is similar: $r^i_{first,final}$ contains information about GP skill μ^i over and above what is contained in $r^i_{first,interim}$, and both follow-on fund size and the expected final return on follow-on funds are determined by μ^i.

F. Additional Funds

Our model assumes that each GP raises at most two funds. In practice, GPs often raise more than two funds over time. Would our theory predict that persistence remains even when comparing returns on, say, funds 2 and 3? To examine this, we construct a simplified version of our model (with risk neutral agents and non-overlapping funds) in which we allow each GP to raise up to three funds. This simplified version (available as an online appendix) demonstrates that performance persistence is present both from fund 1 to 2 and from fund 2 to 3. Intuitively, performance persistence extends to later funds because only a small amount of information asymmetry is required to induce outside investors to withdraw from the market. It does not matter whether the information asymmetry is reduced over time as the performance of later funds is observed. What matters is simply that the information asymmetry remains positive.

G. Optimality of Asymmetric Learning

Learning is valuable whether it happens symmetrically (with incumbent and outside investors learning about GP skill at the same speed) or asymmetrically (with incumbent LPs learning faster than outside investors). It ensures that more skilled GPs receive more capital in follow-on funds and that low-skill GPs exit the industry. This increases the overall value created by the VC industry.

In expectation across first and follow-on funds, LPs earn no rents in utility terms. This implies that the benefits of learning go to the GPs, who thus prefer learning to no learning ex ante.

Can asymmetric learning lead to more efficient fundraising than symmetric learning? In our setting, the answer is yes if LPs find it unattractive to invest in the average GP's first-time fund even at a fund fee of zero:

$$\max_{I_0} E_{\mu^i}\left(e^{-\gamma\left[\frac{1}{2}E\left(C_2^i|\mu^i\right)-\frac{1}{8}\gamma\sigma^2(I_0)^2-\frac{I_0}{2}\right]}\right) < 1.$$

In a risk-neutral setting, this statement would be equivalent to saying that the average NPV of first-time funds at the optimal fund size, averaged across GPs, is negative: $\max_{I_0} E_{\mu^i}\left(E\left(C_2^i|\mu^i\right)-I_0\right) < 0$. Under this condition, a GP would not be able to raise a first-time fund (nor any follow-on funds) if learning was symmetric. However, with asymmetric learning, LPs earn informational rents in follow-on funds, and these may be sufficient to make up for the expected losses on first-time funds.

This will be the case if there is enough dispersion in GP skill.

Effectively, with asymmetric learning, an investment in a first fund gives LPs an *option* to invest in a follow-on fund, and the value of this option increases in uncertainty about GP skill. If the option value equals or exceeds the expected loss on first funds, LPs will invest in first funds despite their negative contribution to expected utility.

The existence of soft information about their skill effectively commits GPs to sharing the value of follow-on funds with their LPs and thus leads to more efficient fund flows. This is also the case in standard models of informational hold-up in the banking literature such as Sharpe (1990), but there investment is inefficient in both periods because interest rates are distorted. No such distortion is present in the VC setting: Fund contracts specify both an investment level (fund size) and the division of the fund's surplus, which, as we have shown, yields first-best fund sizes in each period.

The fact that contracts between GPs and LPs provide exclusive informational rights to incumbent LPs while prohibiting LPs from sharing such information with outsiders is consistent with GPs recognizing that subjecting themselves to informational hold-up may increase their expected utility. Of course, even if this is true ex ante, it is clear that GPs who subsequently learn that they have skill will have an incentive to signal their type to outside investors prior to raising a follow-on fund. In practice, skilled GPs do try to signal their type, but they are unlikely to do so with sufficient precision to eliminate the information asymmetry between incumbent and outside investors. For example, one way that skilled GPs try to signal is by "grandstanding", i.e., taking portfolio firms public earlier than may otherwise be optimal (Gompers (1996)). Grandstanding is unlikely to fully reveal the GP's type, however, since the number of IPOs is unlikely to be fully informative about skill.

Finally, explicit long-term contracts might substitute for incumbent LPs engaging in costly learning. In practice, contracts do not give LPs explicit rights to invest particular amounts at a particular fee should a follow-on fund be raised, suggesting enforcement problems. From a practical standpoint, the problem with long-term contracts is that a court would not be able to enforce either that the correct GPs raise follow-on funds or that the correct follow-on fund size be raised, unless the court itself obtained soft information about GP type. In principle, it would be cost-efficient to have only one party (the court) learn this soft information, rather than having each LP spend resources doing so. In reality, however, such learning would involve court members meeting with GPs and portfolio companies etc. on a regular basis, prior to any potential legal action, which is not consistent with how actual legal processes work.

II. Sample and Data

To examine whether the implications of our model are consistent with empirical patterns observed in the VC industry, we construct a sample of U.S. VC funds obtained from two databases, Thomson Reuters' Venture Economics (VE) and Private Equity Intelligence (PREQIN).[16] As Table 2 shows, our sample contains 2,257 funds raised by 962 VC firms between 1980 and 2002.[17] The number of funds raised per year averages 62 in the 1980s, 106 in the 1990s, and 192 between 2000 and 2002. The average (median) sample fund raised \$111.2 million (\$46.0 million) in nominal dollars. Average fund size increased from \$30.4 million in 1980 to \$46.0 million in 1990, and \$201.4 million in 2000, and then fell to \$130.2 million in 2002 following the ending of the late 1990s tech boom. 39% of sample funds are first-time funds and the average fund sequence number is 2.8.[18] We use fund stage focus as a crude control for differences in risk across funds. 54% of sample funds focus on investing in (usually riskier) early-stage companies.

We are interested in the predictability of fund performance and a VC firm's ability to raise

[16]We define as VC funds all funds listed as focusing on start-up, early-stage, development, late-stage, or expansion investments, as well as those listed as "venture (general)" or "balanced" funds. In cases where VE and PREQIN classify a fund differently, we verify fund type using secondary sources such as *Pratt's Guide*, *CapitalIQ*, *Galante's*, and a web search. We screen out funds of funds, buyout funds, hedge funds, venture leasing funds, evergreen funds (i.e., funds without a predetermined dissolution date), corporate VCs, bank-affiliated funds, SBICs, side funds, and foreign VCs.

[17]VE has the better coverage. Of the 2,257 sample funds, 729 appear in both VE and PREQIN, 37 appear only in PREQIN, and the remaining 1,491 appear only in VE.

[18]While 1980 is our first sample year, not all 1980 vintage funds in the sample are first-time funds. This reflects the fact that our sample contains VC firms founded prior to 1980.

follow-on funds. Since VC funds typically have a ten-year life, we track each sample fund through October 2012, which gives us a minimum of 10 years of performance data, as detailed shortly. We similarly track each of the 962 VC firms through 2012 to see if they raise subsequent funds and thereby manage to stay in business. In addition to the 2,257 funds they raise between 1980 and 2002, sample firms raise another 382 funds between 2003 and October 2012. Still, mortality proves to be high: Using data from *CapitalIQ* combined with fund histories obtained from VE and PREQIN, we find that 661 of the 962 VC firms (68.7%) go out of business between 1980 and 2012.[19] This gives a lower bound on the prevalence of skill in the VC industry of around 1/3, to the extent that GPs will fail to raise a follow-on fund if investors learn that their skill μ^i is less than the break-even level of skill, μ^*.[20] Taking into account that VC firms that survive through 2012 may fail at some point in the future and so are "right-censored", we estimate that the average (median) VC firm fails 14.5 (12) years after founding, having raised 2.7 (2) funds over its lifetime.

A. Interim and Final Performance Data

Our model distinguishes between what incumbent LPs know and what outside investors know at the time they are offered the opportunity to invest in a follow-on fund. To capture this, we distinguish between 'interim' returns, which are observable to all potential investors at the time of fundraising and constitute 'hard' (i.e., verifiable) information based on actual cash flows and audited net asset values, and 'final' or 'ex post' returns which proxy for soft information known to GPs and incumbent LPs at the time of fundraising. As such soft information is unverifiable, it is not known to outside LPs at the time of follow-on fundraising.

We obtain performance data from VE and PREQIN. VC funds are under no obligation to disclose performance data publicly though they share data with their incumbent LPs on a regular basis and with prospective investors whenever they launch a new fund. While these disclosures are

[19]Defunct VC firms are those *CapitalIQ* labels "out of business", "dissolved", "liquidating", "no longer investing", or "reorganizing." We also assume that firms that last raised a fund in 2002 or earlier are defunct as of 2012. Some of these are listed in *CapitalIQ* as having "launched" a fund in, say 2004, but evidently without success.

[20]It is an lower bound because VC firms may also fail for idiosyncratic reasons, such as the death of the GP.

intended to be confidential, data vendors such as VE and PREQIN collect performance data from LPs and/or GPs for dissemination to subscribers, usually in aggregate form.

Our tests focus on disaggregated (fund-by-fund) IRRs, calculated net of fees and so representative of an LP's actual return. A fund's performance varies over its ten-year life as it makes deals, exits portfolio companies, or writes off investments.[21] We extract time-varying interim IRRs from VE and PREQIN, where available, for each year a fund is in operation. These allow us to track performance as it evolves over a fund's life (or more specifically, as it is revealed to incumbent LPs and outside investors over time). We also obtain the final IRR, which records a fund's overall performance from inception to the end of its life. Interim IRRs reflect a mixture of objective cash-on-cash returns in respect of exited investments and changes in the book values of unrealized investments. Final IRRs consist only of audited cash-on-cash returns. Our IRR data cover the period 1980 to 2012. Our interim IRRs thus follow the fund annually over at least 10 years, and our final IRRs are the realized returns after at least 10 years of fund life.

Final IRRs are available for 1,052 of the 2,257 funds (46.6%). The average (median) final IRR for funds raised between 1980 and 2002 is 15.7% (5.6%).[22] There is considerable variation over time in these averages. While 1980s and 1990s funds earned an average annual return of 10.1% and 27.9%, respectively, funds raised in 2000-2002 have lost 2.4% on average per year through 2012.

We have interim IRRs for 15,205 fund-years in respect of 944 individual funds.[23] There are frequently gaps at the start of a fund's life, as IRRs are only defined once a fund has experienced a cash inflow from a sale or has written up an investment, both of which are rare early in a fund's life.[24] There can also be gaps in the middle or towards the end of a fund's life, if both VE and PREQIN encountered difficulty obtaining data for a given fund-year. As a result, we have a

[21] As Ljungqvist and Richardson (2003) show, over a fund's life, performance follows a 'J-curve', in the sense that *cash-on-cash* IRRs (rather than reported interim IRRs) tend to be negative in the first few years as the fund is mainly in investment mode and then turn positive after five or six years as the fund begins to exit its investments through IPOs or M&A transactions.

[22] While the data are thus skewed to the right, winsorizing the data does not materially affect our results.

[23] We have more than 10x944 fund-years because VE and PREQIN report IRRs beyond a fund's 10th anniversary. Usually, IRRs change little after year 10.

[24] For this reason, VE and PREQIN often mark IRRs as 'not meaningful' in the first 2-3 years of a fund's life.

complete record of interim performance for each fund-year for only 547 funds. Figure 3 shows how interim IRRs evolve over the average such fund's life. In its launch year (fund year 0), the average fund reports an IRR of 0.6%, rising to 4.1% in year 1, 8.5% in year 2, 10.1% in year 3, 12.2% in year 4, and 14.2% in year 5, before levelling off at a little under 16% in subsequent years.[25]

B. How Accurately Do Interim IRRs Forecast Final Performance?

Asymmetric learning implies that incumbent LPs have better information about a fund's final return, even before the fund's 10 years are up, than do outside investors, who only observe hard information in the form of interim returns. To test this implication, we use final fund returns as a proxy for the soft information incumbent LPs learn over time by virtue of investing in a GP's fund. In other words, we assume that incumbent LPs can more accurately forecast final returns, even well before the fund's life is over, than can outside investors. If this proxy for incumbent LPs' soft information can predict whether a GP raises a follow-on fund as well as the size and final performance of the follow-on fund, controlling for publicly available hard information contained in interim IRRs at the time of fundraising, then learning is plausibly asymmetric.

As a first step in the analysis, we ask how accurately interim IRRs forecast a fund's final performance and thus how useful hard information may be to outside investors. Figure 4 shows box plots of the distribution of 'forecast errors' (measured as the difference between final and interim IRRs) for each year in a fund's life. Here, we use all 15,205 fund-years for which interim IRR are available. Two patterns emerge. First, the average forecast error is positive in every fund-year, which reflects the pattern seen in Figure 3 of average interim returns rising monotonically before converging on the final IRR. More importantly, the distribution of forecast errors is extremely noisy in the early fund-years and narrows monotonically over time as funds reach the end of their ten-year lives. We can think of the noise in interim IRRs as an upper bound on incumbent LPs' informational advantage over outside investors: If incumbent LPs can predict final IRRs perfectly

[25] Note that there are no apparent performance differences between funds for which we do and do not have interim IRRs: Both return between 15% and 16% a year on average over their lifetimes.

based on their soft information, their forecast errors will be zero. More generally, their forecast errors will be smaller than those of outside investors who only have access to noisy interim IRRs.

C. First and Follow-on Funds

Implications 1 through 4 relate the interim and final performance of a GP's first fund to the likelihood that the GP raises a follow-on fund half way through the life of the first fund, as well as to the size and performance of such a follow-on fund if raised. The key insight of the model is that incumbent LPs can make better follow-on investing decisions than outside investors once they have learned the GP's type. In practice, it is an empirical question whether this learning is complete when the GP raises his second fund; after all, the average (median) second fund is raised only 3.1 (3) years into the first fund's life. At this point in time, the first fund will barely have deployed all its capital and will in most cases not yet have experienced any exits and so arguably is still too immature to have generated much information about the GP's skill. Thus, it is plausible that not much learning has taken place yet when GPs raise their second funds.

Important learning milestones, in practice, are whether the GP managed to find enough deals to deploy all his capital and whether any of the deals could be successfully exited. Thus, information about the GP's true quality likely takes quite a long time to learn. Exactly when incumbent LPs learn the GP's true quality is not observed. With layered funds raised every 3-4 years, it may take until fund 3 or 4 for a sufficient number of investment successes and failures to materialize and hence for the incumbent LPs to learn the GP's true quality. For this reason, our models will flexibly distinguish between first and follow-on funds, rather than between first and second funds only.

D. Prior-fund Performance

We use our performance data to proxy for incumbent and outside investors' information sets as of the year prior to fundraising. To operationalize this, we identify the GP's most recent outstanding fund. Because VC funds rarely have meaningful IRRs in their first 2 years of operation, as mentioned earlier, we require this fund be at least 3 years old. If the most recent fund is less than 3 years old, we skip one vintage and identify the fund prior to that. (This happens in 15% of cases.) We then record the chosen fund's interim and final IRRs. For example, ahead of the GP raising his third fund, we measure the interim IRR of his second fund, if that fund is at least 3 years old, or else the interim IRR of his first fund. In either case, we measure performance as of the year before fundraising.

We have prior-fund interim IRRs for 767 follow-on funds and both interim and final IRRs for 684 follow-on funds. Our performance persistence tests additionally require final IRR data for the follow-on funds themselves. This additional requirement results in samples sizes of 387 and 374 funds when conditioning on interim-only and interim-and-final IRRs, respectively.

Note that we use the performance of only the *immediately prior* fund to measure the hard information available to investors. In principle, the performance of older funds, if any, could also contribute to investors' information set. However, if performance is indeed persistent, the return on the immediately prior fund will be a sufficient statistic for the GP's prior funds. As we will show, this is indeed the case; conditioning on the performance of older funds does not affect our results.

III. Empirical Analysis

The focus of our empirical analysis is on the role of asymmetric learning and soft information in explaining performance persistence and future fundraising in VC. We first replicate the motivating fact of our paper, namely that VC fund performance is persistent. We then ask if privately available soft information can predict performance and fundraising over and above publicly available hard information and find that it can. Finally, we discuss possible alternative explanations for persistence.

A. Persistence, Learning, and Soft Information

A.1. Performance Persistence

We begin by replicating Kaplan and Schoar's (2005) persistence test in our larger dataset. In column 1 of Table 3, we regress a fund's ex post IRR on log fund size, the ex post IRR of the VC firm's previous fund, and vintage-year effects. Standard errors are clustered by VC firm. Like Kaplan and Schoar, we find that fund performance increases with fund size and prior-fund performance ($p < 0.001$).

One concern regarding the persistence result is selection bias: Not every VC fund reports an IRR, and it is possible that those that do are those that experience persistent good performance. To explore the extent of this bias, we estimate a persistence regression with exit rates as the dependent variable instead of IRRs. Hochberg, Ljungqvist, and Lu (2007) define exit rates as the fraction of a fund's investments that were exited through an IPO or an M&A transaction over the course of the fund's ten-year life. Exit rates can thus be computed for all funds. As the estimates in column 2 show, we continue to find strong evidence of persistence using this alternative performance measure.

A.2. What Type of Information Predicts Returns?

According to Implication 1a, a high interim return on one fund should predict a high final return on the GP's next fund. We test this in column 3 of Table 3. The results strongly support the prediction. The coefficient on the prior-fund interim IRR, measured as of the year before the current fund was raised, is positive and highly statistically significant ($p = 0.004$).

Implication 1b states that interim returns should be informative even after adjusting for idiosyncratic risk. Since VC funds are not traded, traditional asset pricing proxies for idiosyncratic risk are not available. Instead, we follow Kaplan and Schoar (2005) and include a dummy variable that equals 1 for funds classified as investing in early-stage companies as a crude control for differences in risk-taking across funds. Figure 5 shows kernel density estimates for the final returns of early-stage and late-stage funds. The distribution of early-stage fund returns is considerably more fat-tailed, consistent with the interpretation that early-stage funds take more risk. A formal Kolmogorov-Smirnov test confirms that the two distributions are significantly different from each other ($p = 0.002$). In column 4 of Table 3, we see that average returns among early-stage funds are 6.9 percentage points higher than among late-stage funds ($p = 0.075$). Controlling for risk using this proxy does not, however, change our conclusion that interim returns significantly predict the future returns of follow-on funds. (Indeed, the point estimates are nearly identical in columns 3 and 4.) This supports Implication 1b.

If learning is indeed asymmetric, as our model assumes, soft information about GP skill should help predict LP returns in the GP's next fund, over and above the hard information available to outside investors at the time the next fund is raised. This is Implication 4 of the model. This implication, along with those relating future fundraising to soft information, potentially allows us to discriminate between symmetric and asymmetric learning and so to test our model.

In column 5, we run a horse race between the prior fund's interim IRR (measured as of the year-end prior to the year the GP raised the current fund) and its future ex post return. As predicted, both correlate positively and statistically significantly with the next fund's final IRR. The point estimate is five times larger, and less noisy, for ex post than for interim IRRs. This suggests that ex post IRRs contain more information about future performance than do interim IRRs.[26] A look at the regression R^2 confirms this. Compared to column 4, adding ex post returns substantially increases the adjusted R^2, from 16.7% to 23%. Thus, the ex post IRR of a GP's previous fund appears to be highly informative about the performance of the GP's next fund. This pattern is consistent with the informational assumptions of our model: Information not yet publicly known at the time of fundraising (i.e., ex-post IRRs) predicts returns on follow-on funds over and above hard information known at the time of fundraising (i.e., interim IRRs).

One potential confound that could spuriously lead to greater persistence with respect to ex

[26] This remains the case if we condition not only on the prior fund's interim IRR but on hard information relating to the performance of *all* the funds the GP managed before. For example, the coefficient on a variable capturing the highest return the GP ever achieved before the prior fund is insignificant ($p = 0.381$), and including this variable has next to no effect on the point estimates of the prior fund's interim and final IRRs.

post IRRs than to interim IRRs is the fact (documented in Figures 3 and 4) that average interim returns rise monotonically over the life of a fund before converging on the final IRR. Suppose that low-skilled GPs struggle to raise follow-on funds and so tend to raise funds when their prior fund is older. Then, given the patterns in Figures 3 and 4, they will tend to report higher interim IRRs than do highly-skilled GPs at the time of fundraising. If low-skilled GPs earn low returns on their follow-on funds, this will then attenuate the predictive power of interim IRRs relative to ex post IRRs. A simple way to account for this is to condition on the age of the prior fund. Doing so has virtually no effect on our findings (see column 6), suggesting that this potential confound is not a serious concern in the data.

B. Effect of Learning on Fund-Raising

The results discussed in the previous section support Implications 1a and 1b, which hold even if learning is symmetric. Implication 4, on the other hand, is true only if learning is asymmetric and the fact that it appears to hold in the data suggests that informational hold-up may be the underlying cause of performance persistence. We can shed further light on this by relating the likelihood that a GP raises a follow-on fund, and the size of that follow-on fund if raised, to the information available to incumbent LPs and outside investors, respectively. Implications 2a and 3a state that publicly available 'hard' information should predict future fundraising, as investors use this information to update their priors about the GP's type. But if learning is asymmetric, as our model assumes, then our proxy for incumbent LPs' 'soft' information should predict future fundraising over and above the publicly available information (Implications 2b and 3b). This distinction allows us to discriminate between symmetric and asymmetric learning in the data.

B.1. Probability of Future Fundraising

To test Implication 2a, we estimate a Cox hazard model with time-varying covariates, which can capture how changes in reported interim IRRs affect the probability that a VC firm raises a new fund the following year. Column 1 of Table 4 reports the coefficient estimates. Controlling for the fact that VC firms with larger funds are more likely to raise another fund, we find that higher interim returns on the previous fund significantly increase the hazard of raising a new fund ($p < 0.001$). A one standard deviation increase in the prior fund's interim IRR as of year $t-1$ (39.2%) is associated with an 11.2 percentage-point increase in the likelihood of raising a follow-on fund in year t. This supports Implication 2a.

Column 2 additionally conditions on the prior fund's final IRR, which will not be publicly known until, on average, 7 years later. The results strongly support Implication 2b and thus *asymmetric* learning. A one-standard-deviation higher ex post IRR on the previous fund increases the likelihood that the GP will raise a follow-on fund in year t by 8.3 percentage points ($p = 0.001$). The corresponding influence of publicly available interim IRRs, on the other hand, is halved compared to column 1 ($p = 0.024$).

B.2. Size of Follow-on Fund

According to Implication 3a, the size of a follow-on fund, if raised, increases in the prior fund's interim return. To test this, we need to allow for the possibility that a poorly performing VC firm will be unable to raise a follow-on fund of any size. (Recall that 661 of the 962 VC firms fail to raise follow-on funds over our sample period and so go out of business.) This means that the dependent variable is left-censored and needs to be modeled using a Tobit estimator. The dependent variable then equals the log fund size if the firm raises a follow-on fund and zero if it does not.

The results are presented in column 3 of Table 4. As predicted, we find that good interim performance for the GP's previous fund allows the GP to raise a larger follow-on fund. A one-standard deviation increase in the previous fund's interim IRR is associated with a 145% or $49.5 million increase in fund size, from the unconditional mean in the estimation sample of $34.2 million ($p < 0.001$). This supports Implication 3a.

When we additionally condition on the prior fund's final IRR, which outside investors do not observe, we find evidence consistent with Implication 3b and so with asymmetric learning. A one-standard-deviation increase in the ex post IRR on the GP's previous fund leads to an additional boost in follow-on fund size of 25.4% or \$9.2 million ($p = 0.042$ in column 4).

C. Alternative Explanations

The evidence in Table 3 shows that future fund returns can be predicted using prior funds' future ex post IRRs, which will not be known until some years after fundraising, even controlling for publicly available information in the form of prior funds' interim returns. Table 4 then shows that prior funds' future ex post IRRs can predict both whether the GP raises a follow-on fund and if so, how large the follow-on fund will be. A plausible explanation for these findings is that ex post IRRs correlate with incumbent LPs' private (soft) information. In other words, incumbent LPs appear to know something that is not captured by publicly available interim performance measures and which allows them to make reinvestment decisions that resemble the return-chasing behavior seen in mutual funds—except that the returns being chased are not yet publicly known.

We are not aware of any alternative explanation for performance persistence that would predict a differential role for 'soft' over 'hard' information or that could account for the additional fundraising patterns we see in the data. Nonetheless, it is worth considering two potential alternative explanations that have been advanced for Kaplan and Schoar's (2005) finding that performance persists in VC.

The main alternative explanation is due to Glode and Green (2011). Set in the context of hedge funds, their model emphasizes asymmetric learning about the nature of the GP's strategy. This allows incumbent LPs to threaten to 'steal' the strategy (i.e., reveal it to another GP) and thereby extract part of the follow-on fund's surplus, generating persistence. Our model instead formalizes the informational hold-up resulting from asymmetric learning about skill rather than strategy: How good is the VC at identifying promising start-ups and screening out losers and how much value does he add to his investments through strategic advice, help in recruiting talent, and access to his rolodex of contacts? If these are skills that incumbent LPs can 'steal', Glode and Green's model applies. If instead knowledge of these skills enables incumbent LPs to hold the GP up, our model applies.

Da Rin and Phalippou's (2011) survey, discussed in the introduction, attempts to test which of these two models better applies in the VC setting. As our Table 1 shows, only 13.1% of LPs in the survey agreed with the following statement: "If the GP didn't allow me to reinvest, I could replicate their strategy (myself or in cooperation with another GP)." This suggests that stealing the investment strategy is less of a concern in the VC setting. In contrast, 72.1% of these LPs agreed with the statement, "If I didn't re-invest, other investors would be suspicious and would not invest," supporting the informational hold-up story.

An informal argument popular with industry professionals for why GPs do not increase their fees, eliminating persistence, is that GPs cede a share of their rents to LPs to ensure they can raise funds even in bad times. This argument does not, however, predict persistence in and of itself: If every GP cedes a constant amount, there is no persistence. To obtain persistence, skilled GPs would have to offer LPs a higher return on all their funds while less skilled GPs offer LPs a lower return on all of theirs. [This could occur if, for instance, the following took place: (1) Less skilled GPs raised funds only in good times – defined as times where investors require lower expected returns to invest – and skilled GPs raised funds in both good and bad times, and (2) each GP offered an expected return on all his funds equal to the average return investors require across the funds he raises.[27] We would then observe what looks like performance persistence, but it would be a result of differences in required discount rates in different periods.

It is not obvious that a suitably augmented practitioner story would have anything to say about the predictive power of soft information in the form of final returns, over and above publicly observable interim returns. Still, it is worth attempting to empirically distinguish it from informational hold-up as follows. The practitioner story implies that we should not see persistence in the subset

[27]Exactly why such expected return smoothing would be used is not clear in this story.

of skilled GPs (those who are able to raise funds in both good and bad times). In Table 5, we thus restrict our sample to GPs that raise funds in both good and bad times, using four different classifications of "good" and "bad" periods. We observe strong performance persistence in all four cases, which is hard to reconcile with the practitioner story.

IV. Discussion and Conclusion

Performance in the VC market appears persistent, suggesting (some) VCs have skill. But why then do successful VCs not eliminate excess demand for their next funds by raising their fees? We propose a model of learning and informational hold-up that can explain performance persistence in the VC market. We argue that persistence requires that the LP market is perfectly competitive when a GP raises his first fund and that his investors subsequently gain market power. We propose that the source of their market power is asymmetric learning: Investing in a fund gives an LP the opportunity to collect soft information about the GP's skill, while outside investors can observe only hard information such as realized returns. Thus, incumbent LPs have an informational advantage when the GP raises his next fund. This imposes a winner's curse on outside investors—the better-informed incumbent LPs will outbid them whenever the GP has skill—and enables incumbent LPs to hold the GP up when he next raises a fund. Performance is persistent because the hold-up problem prevents the GP from raising his fees to the point where investors simply break even.

The driving force of our model is initial uncertainty about GP skill which is resolved more quickly among incumbent LPs than among potential outside investors. Thus, the information sets of incumbent LPs and outside investors diverge over time. According to our model, the information held by the better-informed incumbent LPs predicts not only the performance of the GP's next fund (since it is informative about his skill) but also whether the GP can raise a follow-on fund and, if so, of what size. We verify these predictions with one of the most comprehensive datasets on U.S. VC funds assembled to date. Though the inference is necessarily indirect, these patterns point to incumbent LPs obtaining private information about GP skill and so are at least consistent with asymmetric learning. Survey evidence that directly addresses the hold-up story provides additional supportive evidence for our theory.

References

Ang, A., D. Papanikolaou, and M. M. Westerfield, 2012, "Portfolio Choice with Illiquid Assets," Unpublished Working Paper, Columbia University.

Anton, J., and D. Yao, 1989, "Split Awards, Procurement, and Innovation," *RAND Journal of Economics* 20, 538-552.

Berk, J., and R. Green, 2004, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy* 112, 1269-1295.

Binmore, K., A. Rubinstein, and A. Wolinsky, 1986, "The Nash Bargaining Solution in Economic Modelling," *RAND Journal of Economics* 17, 176-188.

Binmore, K., M. J. Osborne, and A. Rubinstein, 1992, "Non-Cooperative Models of Bargaining," Handbook of Game Theory, Volume 1, Edited by R. J. Aumann and S. Hart.

Blaydon, C., and M. Horvath, 2002, "What's a Company Worth? Depends Which GP You Ask," *Venture Capital Journal (May)*, 40-41.

Chung, J.W., B.A. Sensoy, L.H. Stern, and M.S. Weisbach, 2010, "Incentives of Private Equity General Partners from Future Fundraising," Unpublished working paper, Ohio State University.

Cochrane, J., 2005, "The Risk and Return of Venture Capital," *Journal of Financial Economics* 75, 3-52.

Cornelli, F., Z. Kominek, and A. Ljungqvist, 2012, "Monitoring Managers: Does It Matter?," *Journal of Finance*, forthcoming.

Da Rin, M., and L. Phalippou, 2011, "On the Heterogeneity of Private Equity Investors," Unpublished working paper, Oxford University.

Glode, V., and R. Green, 2011, "Information Spillovers and Performance Persistence for Hedge Funds," *Journal of Financial Economics* 101, 1-17.

Gompers, P.A., 1996, "Grandstanding in the Venture Capital Industry," *Journal of Financial Economics* 43, 133-156.

Hochberg, Y. V., Ljungqvist A., and Y. Lu, 2007, "Whom You Know Matters: Venture Capital Networks and Investment Performance," *Journal of Finance* 62, 251-301.

Jones, C., and M. Rhodes-Kropf, 2003, "The Price of Diversifiable Risk in Venture Capital and Private Equity," Unpublished working paper, Columbia University.

Kaplan, S.N., and A. Schoar, 2005, "Private Equity Performance: Returns, Persistence and Capital Flows," *Journal of Finance* 60, 1791-1823.

Korteweg, A., and M. Sørensen, 2010, "Risk and Return Characteristics of Venture Capital-Backed Entrepreneurial Companies," *Review of Financial Studies* 23, 3738-3772.

Lerner, J., and A. Schoar, 2004, "The Illiquidity Puzzle: Evidence from Private Equity Partnerships," *Journal of Financial Economics* 72, 3-40.

Lerner, J., A. Schoar, and W. Wongsunwai, 2007, "Smart Institutions, Foolish Choices?: The Limited Partner Performance Puzzle," *Journal of Finance* 62, 731-764.

Ljungqvist, A., and M. Richardson, 2003, "The Cash Flow, Return, and Risk Characteristics of Private Equity," NBER Working Paper No. 9454.

Ljungqvist, A., M. Richardson, and D. Wolfenzon, 2007, "The Investment Behavior of Buyout Fund Managers," Unpublished working paper, NYU.

Lopez de Silanes, F., L. Phalippou, and O. Gottschalg, 2010, "Giants at the Gate: Diseconomies of Scale in Private Equity," Unpublished working paper, University of Amsterdam.

Malkiel, B.G., 1995, "Returns from Investing in Equity Mutual Funds 1971-1991," *Journal of Finance* 50, 549-572.

Pastor, L., and P. Veronesi, 2009, "Learning in Financial Markets," *Annual Review of Financial Economics* 1, 361-381.

Quigley, J., and S. Woodward, 2003, "An Index for Venture Capital," Unpublished working paper, University of California, Berkeley.

Rajan, R., 1992, "Insiders and Outsiders: The Choice between Informed and Arm's Length Debt," *Journal of Finance* 47, 1367-1400.

Rubinstein, A., 1982, "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50, 97-109.

Sharpe, S.A., 1990, "Asymmetric Information, Bank Lending and Implicit Contracts: A Stylized Model of Customer Relationships," *Journal of Finance* 45, 1069-1087.

Sorensen, M., N. Wang, and J. Yang, 2012, "Valuing Private Equity," Unpublished Working Paper, Columbia University.

Von Thadden, E.-L., 2004, "Asymmetric Information, Bank Lending and Implicit Contracts: The Winner's Curse," *Finance Research Letters* 1, 11-23.

A Appendix: Derivations and Proofs

Derivation of optimal fund sizes for the case of two and one LPs, respectively:

Two LPs: $I^i_{1,split}$ solves

$$\max_{I_1} E\left(U^{GP}|\mu^i\right) + 2E\left(U^{LP}|\mu^i\right) \tag{A.1}$$

where

$$E\left(U^{GP}|\mu^i\right) = 1 - e^{-\gamma W_3^{GP}} = 1 - e^{-\gamma\left[W_0^{GP}+2M_0+2M_1\right]} \tag{A.2}$$

and

$$\begin{aligned} E\left(U^{LP}|\mu^i\right) &= 1 - E\left(e^{-\gamma W_3^{LP}}|\mu^i\right) \\ &= 1 - E\left(e^{-\gamma\left[W_0^{LP}+\frac{1}{2}(A_2\ln(1+I_0)-I_0)-M_0+\frac{1}{2}(A_3\ln(1+I_1)-I_1)-M_1\right]}|\mu^i\right) \\ &= 1 - e^{-\gamma\left[W_0^{LP}-M_0-M_1\right]} E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]}|\mu^i\right) E\left(e^{-\gamma\frac{1}{2}[A_3\ln(1+I_1)-I_1]}|\mu^i\right) \end{aligned} \tag{A.3}$$

We exploit the fact that A_2 and A_3 are independent, conditional on μ^i. Furthermore, since A_3 is normally distributed, and since $V\left(A_3|\mu^i\right) = \frac{\sigma^2(I_1)^2}{[ln(1+I_1)]^2}$, we obtain $E\left(e^{-\gamma\frac{1}{2}[A_3\ln(1+I_1)-I_1]}|\mu^i\right) = e^{-\gamma\frac{1}{2}\left(E\left(A_3|\mu^i\right)\ln(1+I_1)-I_1\right)+\frac{1}{8}\gamma^2\sigma^2(I_1)^2}$.

Maximizing the joint surplus thus implies solving:

$$\max_{I_1}\left(E\left(A_3|\mu^i\right)\ln\left(1+I_1\right) - I_1\right) - \gamma\frac{1}{4}\sigma^2\left(I_1\right)^2 \tag{A.4}$$

which has solution:

$$I^i_{1,split} = \frac{E\left(A_3|\mu^i\right)}{1+\gamma\frac{1}{2}\sigma^2 I_{1,split}\left(\mu^i\right)} - 1 \iff \tag{A.5}$$

$$I^i_{1,split} = \frac{-\left(1+\gamma\frac{1}{2}\sigma^2\right)+\sqrt{\left(1+\gamma\frac{1}{2}\sigma^2\right)^2 - 2\gamma\sigma^2\left[1-E\left(A_3|\mu^i\right)\right]}}{\gamma\sigma^2}. \tag{A.6}$$

One LP: If only one LP invests in the follow-on fund, maximizing the joint surplus implies

$$\max_{I_1} E\left(U^{GP}|\mu^i\right) + E\left(U^{LP}|\mu^i\right). \tag{A.7}$$

Here, the term $E\left(e^{-\gamma\frac{1}{2}[A_3\ln(1+I_1)-I_1]}|\mu^i\right)$ in the LP's expected utility changes to $E\left(e^{-\gamma[A_3\ln(1+I_1)-I_1]}|\mu^i\right)$ compared to the two-LP scenario. This implies maximizing $E\left(A_3|\mu^i\right)\ln(1+I_1)-I_1-\gamma\frac{1}{2}\sigma^2(I_1)^2$, which results in a smaller joint-surplus-maximizing fund size of

$$I^i_{1,sole} = \frac{E\left(A_3|\mu^i\right)}{1+\gamma\sigma^2 I^i_{1,sole}} - 1 \Longleftrightarrow \tag{A.8}$$

$$I^i_{1,sole} = \frac{-\left(1+\gamma\sigma^2\right)+\sqrt{\left(1+\gamma\sigma^2\right)^2-4\gamma\sigma^2\left[1-E\left(A_3|\mu^i\right)\right]}}{2\gamma\sigma^2}. \tag{A.9}$$

It follows that both $I^i_{1,split}$ and $I^i_{1,sole}$ equal zero for $E\left(A_3|\mu^i\right)=1$. Since $E\left(A_3|\mu^i\right)=a+2\mu^i$, this implies that the cutoff GP type for a follow-on fund generating no joint surplus is given by $a+2\mu^i=1 \Longleftrightarrow \mu^i=\frac{1-a}{2}$. We denote this value of μ^i by μ^*.

Note that the LPs' risk-adjusted cash flow before fees in the split case,

$$Y=\frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right]-\frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2, \tag{A.10}$$

is zero for $E\left(A_3|\mu^i\right)=1$ (i.e., for $\mu^i=\mu^*$) and positive for $E\left(A_3|\mu^i\right)>1$. This follows from

$$\frac{dY}{dE\left(A_3|\mu^i\right)}=\frac{1}{2}\ln\left(1+I^i_{1,split}\right)+\left[\frac{1}{2}\left(\frac{E\left(A_3|\mu^i\right)}{1+I^i_{1,split}}-1\right)-\gamma\frac{1}{4}\sigma^2 I^i_{1,split}\right]\frac{dI^i_{1,split}}{dE\left(A_3|\mu^i\right)} \tag{A.11}$$

where $\frac{1}{2}\ln\left(1+I^i_{1,split}\right)>0$ and $\left[\frac{1}{2}\left(\frac{E\left(A_3|\mu^i\right)}{1+I^i_{1,split}}-1\right)-\gamma\frac{1}{4}\sigma^2 I^i_{1,split}\right]>0$ for $E\left(A_3|\mu^i\right)>1$, and $\frac{dI^i_{1,split}}{dE\left(A_3|\mu^i\right)}\geq 0$ for all values of $E\left(A_3|\mu^i\right)$.

Proof of Proposition 1

(a) Part (a) is true for any value of p. Consider an offer $\left[(I^{GP}_{1,split}/2, M^{GP}_{1,split}),(I^{GP}_{1,sole}, 2M^{GP}_{1,sole})\right]$ with fund sizes $I^{GP}_{1,split}$ and $I^{GP}_{1,sole}$ that are different from $I^i_{1,split}$ and $I^i_{1,sole}$. By definition, $I^i_{1,split}$ and $I^i_{1,sole}$ are the joint-surplus-maximizing fund sizes, and so the GP can always make himself better off by changing the proposed fund sizes to $I^i_{1,split}$ and $I^i_{1,sole}$ and adjusting the proposed fees to make the LPs equally happy. A similar argument applies to offers made by the LPs.

(b) The LPs' expected utility from $(I^i_{1,split}/2, M^*_1)$ is

$$E\left(U^{LP}_{split}|\mu^i\right)=1-e^{-\gamma\left[W^{LP}_0+Z^{LP}_2\right]}e^{-\gamma\frac{1}{2}\left(E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)+\frac{1}{2}\gamma^2\left(\frac{1}{2}\right)^2\sigma^2\left(I^i_{1,split}\right)^2+\gamma M^*_1}$$

where Z_2^{LP} is defined as in the main text. The LPs' expected utility from $(I^i_{1,sole}, 2M_1^*)$ is:

$$E\left(U^{LP}_{sole}|\mu^i\right) = 1 - e^{-\gamma\left[W_0^{LP}+Z_2^{LP}\right]}e^{-\gamma\left(E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,sole}\right)-I^i_{1,sole}\right)+\frac{1}{2}\gamma^2\sigma^2\left(I^i_{1,sole}\right)^2+\gamma 2M_1^*}.$$

It follows that $E\left(U^{LP}_{split}|\mu^i\right) > E\left(U^{LP}_{sole}|\mu^i\right)$ iff the condition stated in Proposition 1 holds.

(c) With two LPs investing, the fees $M^{LP,*}_{1,split}$ and $M^{GP,*}_{1,split}$ that make the GP and the LPs indifferent between accepting the other party's split offer now or having their own split offer accepted in the next offer round solve the following two equations. For any p, the GP's indifference condition is

$$\begin{aligned}
1 - e^{-\gamma\left[W_0^{GP}+2M_0+2M^{LP,*}_{1,split}\right]} &= p\left[1 - e^{-\gamma\left[W_0^{GP}+2M_0\right]}\right] + (1-p)\left[1 - e^{-\gamma\left[W_0^{GP}+2M_0+2M^{GP,*}_{1,split}\right]}\right] \Longleftrightarrow \\
e^{-\gamma 2M^{LP,*}_{1,split}} &= p + (1-p)\, e^{-\gamma 2M^{GP,*}_{1,split}} \Longleftrightarrow \\
e^{\gamma 2M^{GP,*}_{1,split}} &= \frac{1-p}{e^{-\gamma 2M^{LP,*}_{1,split}} - p}
\end{aligned}$$

Each LP's indifference condition is

$$\begin{aligned}
& 1 - e^{-\gamma W_0^{LP}} E\left(e^{-\gamma Z_2^{LP}}|\mu^i\right) E\left(e^{-\gamma\left[\frac{1}{2}\left(A_3\ln(1+I_1)-I_1\right)-M^{GP,*}_{1,split}\right]}|\mu^i\right) \\
= \; & p\left[1 - e^{-\gamma W_0^{LP}} E\left(e^{-\gamma Z_2^{LP}}|\mu^i\right)\right] \\
& + (1-p)\left[1 - e^{-\gamma W_0^{LP}} E\left(e^{-\gamma Z_2^{LP}}|\mu^i\right) E\left(e^{-\gamma\left[\frac{1}{2}\left(A_3\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)-M^{LP,*}_{1,split}\right]}|\mu^i\right)\right]
\end{aligned}$$

$\Longleftrightarrow$

$$\begin{aligned}
& E\left(e^{-\gamma\left[\frac{1}{2}\left(A_3\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)-M^{GP,*}_{1,split}\right]}|\mu^i\right) \\
&= p + (1-p)\, E\left(e^{-\gamma\left[\frac{1}{2}\left(A_3\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)-M^{LP,*}_{1,split}\right]}|\mu^i\right) \Longleftrightarrow \\
e^{-b+\gamma M^{GP,*}_{1,split}} &= p + (1-p)\, e^{-b+\gamma M^{LP,*}_{1,split}}
\end{aligned}$$

where $Z_2^{LP} = \frac{1}{2}\left(A_2\ln\left(1+I_0\right)-I_0\right) - M_0$ (the LP's payoff from the first-time fund) and

$$b\left(\mu^i\right) = \gamma\frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right) - I^i_{1,split}\right] - \frac{1}{8}\gamma^2\sigma^2\left(I^i_{1,split}\right)^2.$$

Combining the two indifference conditions implies

$$e^{-b}\left(\frac{1-p}{e^{-\gamma 2M_{1,split}^{LP,*}}-p}\right)^{1/2}=p+(1-p)\,e^{-b+\gamma M_{1,split}^{LP,*}}.$$

Denote $e^{-\gamma M_{1,split}^{GP,*}}$ by x. Then the above can be rewritten as

$$\begin{aligned}
e^{-b}\left(\frac{1-p}{x^2-p}\right)^{1/2} &= p+(1-p)\,e^{-b}\frac{1}{x}\\
e^{-2b}\left(\frac{1-p}{x^2-p}\right) &= p^2+(1-p)^2\,e^{-2b}\frac{1}{x^2}+2p\,(1-p)\,e^{-b}\frac{1}{x}\\
e^{-2b}\,(1-p) &= p^2\left(x^2-p\right)+(1-p)^2\,e^{-2b}\left(1-\frac{p}{x^2}\right)+2p\,(1-p)\,e^{-b}\frac{1}{x}\left(x^2-p\right)\\
e^{-2b}\,(1-p)\,x^2 &= p^2\left(x^4-px^2\right)+(1-p)^2\,e^{-2b}\left(x^2-p\right)+2p\,(1-p)\,e^{-b}\left(x^3-px\right)\\
(1-p)\,x^2 &= p^2e^{2b}\left(x^4-px^2\right)+(1-p)^2\left(x^2-p\right)+2p\,(1-p)\,e^{b}\left(x^3-px\right)\\
0 &= p^2e^{2b}x^4+2p\,(1-p)\,e^{b}x^3+\left[p\,(p-1)-p^3e^{2b}\right]x^2-2p\,(1-p)\,e^{b}px-p\,(1-p)^2\\
0 &= pe^{2b}x^4+2\,(1-p)\,e^{b}x^3+\left[(p-1)-p^2e^{2b}\right]x^2-2\,(1-p)\,e^{b}px-(1-p)^2
\end{aligned}$$

$$0=pe^{2b}x^4+2e^b\,(1-p)\,x^3+\left(-1+p-p^2e^{2b}\right)x^2-2pe^b\,(1-p)\,x-(1-p)^2\,.$$

This is a continuous function of p. Thus, as p goes to zero, x solves

$$0=2e^bx^3-x^2-1$$

which has the solution:

$$\begin{aligned}
x &= -\frac{-1}{6e^b}-\frac{1}{6e^b}\left(\frac{1}{2}\left[-2-27*4e^{2b}+\sqrt{[-2-27*4e^{2b}]^2-4}\right]\right)^{1/3}\\
&\quad -\frac{1}{6e^b}\left(\frac{1}{2}\left[-2-27*4e^{2b}-\sqrt{[-2-27*4e^{2b}]^2-4}\right]\right)^{1/3}.
\end{aligned}$$

Given this solution for x, $M_{1,split}^{GP,*}=\frac{-\ln(x)}{\gamma}$ and $M_{1,split}^{LP,*}$ equals $M_{1,split}^{GP,*}$ by the GP's indifference condition when $p\to 0$. We denote this common value of $M_{1,split}^{GP,*}$ and $M_{1,split}^{LP,*}$ by M_1^*.

(d) We need to show that each party's strategy is an optimal response to the strategies of the other two parties.

First consider the GP. The GP cannot do better by increasing $M^{GP}_{1,split}$ or $M^{GP}_{1,sole}$ above M^*_1 since LPs will reject all such offers. Furthermore, under the proposed strategies, LPs accept a fee of M^*_1, and therefore the GP has no incentive to suggest a lower fee.

Next consider LP$_a$ (similar arguments apply to LP$_b$). LP$_a$ cannot do better by decreasing $M^{LP_a}_{1,split}$ or $M^{GP}_{1,sole}$ below M^*_1 since the GP's strategy rejects all such offers. Note that, in this respect, it is important that each LP offers a sole fee as high as $2M^*_1\left(\mu^i\right)$. If they offered less, the other LP could reduce his offered split fee below $M^*_1\left(\mu^i\right)$ and still be asked to invest. But if both LPs lowered their offered split fees below $M^*_1\left(\mu^i\right)$, the GP would pick one of the LPs' sole offers. This would be strictly worse for both LPs (by point (b) for the investing LP and because the non-investing LP would earn nothing from the follow-on fund).

Furthermore, under the proposed strategies, the GP accepts offers with a fee of M^*_1, so LP$_a$ has no incentive to increase $M^{LP_a}_{1,split}$ above M^*_1. If LP$_a$ did so, then the GP would accept both LPs' split offers (earning fees of $M^{LP_a}_{1,split} + M^*_1 > 2M^*_1$), and therefore LP$_a$ would end up with the same investment of $I^i_{1,split}/2$ but would pay a higher fee. In addition, LP$_a$ has no incentive to increase $M^{LP_a}_{1,sole}$ above $2M^*_1$. Doing so would result in the GP accepting LP$_a$'s sole offer, but since LP$_a$ has higher utility from $(I^i_{1,split}/2, M^*_1)$ than $(I^i_{1,sole}, 2M^*_1)$, he will also have higher utility from $(I^i_{1,split}/2, M^*_1)$ than $(I^i_{1,sole}, M^{LP_a}_{1,sole})$ with $M^{LP_a}_{1,sole} > 2M^*_1$.

Proof of Corollary 2:

Recall that for $\mu^i > \mu^*$ (i.e., for $E\left(A_3|\mu^i\right) > 1$), the LPs' risk-adjusted cash flow in a follow-on fund before fees (for the case of two investors),

$$\frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right) - I^i_{1,split}\right] - \frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2, \tag{A.12}$$

is positive. Note that this expression is simply $b\left(\mu^i\right)/\gamma$. The LPs' risk-adjusted cash flow after fees is therefore

$$\begin{aligned} &\frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right) - I^i_{1,split}\right] - \frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2 - M^*_1\left(\mu^i\right) \qquad \text{(A.13)} \\ = &\frac{b\left(\mu^i\right)}{\gamma} - M^*_1\left(\mu^i\right) = \frac{b\left(\mu^i\right)}{\gamma}\left[1 - \frac{\gamma M^*_1\left(\mu^i\right)}{b\left(\mu^i\right)}\right] = \frac{b\left(\mu^i\right)}{\gamma}\left[1 + \frac{\ln\left(x\left(\mu^i\right)\right)}{b\left(\mu^i\right)}\right]. \end{aligned}$$

Using the expression for x as a function of b from the proof of Proposition 1, it is easy to verify (by plotting $\frac{\ln(x(\mu^i))}{b(\mu^i)}$ as a function of $b\left(\mu^i\right)$) that $\frac{\gamma M_1^*(\mu^i)}{b(\mu^i)}$ goes monotonically from 1/2 to 1/3 as $b\left(\mu^i\right)$ goes from 0 to ∞, implying that M_1^* goes from $\frac{1}{2}\frac{b(\mu^i)}{\gamma}$ to $\frac{1}{3}\frac{b(\mu^i)}{\gamma}$ as $b\left(\mu^i\right)$ goes from 0 to ∞. In sum, the risk-adjusted cash flow after fees to each LP, $\frac{b(\mu^i)}{\gamma} - M_1^*\left(\mu^i\right)$, is always positive for $\mu^i > \mu^*$ and

$$\frac{b\left(\mu^i\right)}{\gamma} - M_1^*\left(\mu^i\right) \rightarrow \begin{cases} \frac{1}{2}\frac{b(\mu^i)}{\gamma} \text{ as } b\left(\mu^i\right) \rightarrow 0 \\ \frac{2}{3}\frac{b(\mu^i)}{\gamma} \text{ as } b\left(\mu^i\right) \rightarrow \infty \end{cases}. \tag{A.14}$$

Since $b\left(\mu^i\right)$ increases in μ^i, the above implies that $\frac{b(\mu^i)}{\gamma} - M_1^*\left(\mu^i\right)$ also increases in μ^i. In other words, the risk-adjusted cash flow before fees $\frac{b(\mu^i)}{\gamma}$ increases in μ^i, and the fraction of this that is paid in fees decreases in μ^i from 1/2 (for the marginal follow-on fund with $b\left(\mu^i\right)$ just above zero, i.e., μ^i just above μ^*) to 1/3 (as $\mu^i \rightarrow \infty$ and thus $b\left(\mu^i\right) \rightarrow \infty$). Thus, the LPs' risk-adjusted cash flow after fees also increases in μ^i.

Proof of Implication 1

We prove Implication 1b. Since the risk-adjustment is increasing in μ^i, Implication 1a follows immediately from Implication 1b. Proving Implication 1b requires us to prove that the expectation of

$$1 + r^i_{follow-on,final,risk-adj} = \frac{\frac{1}{2}C_3^i - M_1^*\left(\mu^i\right) - \frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2}{\frac{1}{2}I^i_{1,split}}$$

conditional on $r^i_{first,interim}$ is increasing in $r^i_{first,interim}$.

Step 1: We start by showing that $E\left(r^i_{follow-on,final,risk-adj}|\mu^i, \mu > \mu^i > \mu^*\right)$ is positive and increasing in μ^i.

$$\begin{aligned} E\left(r^i_{follow-on,final,risk-adj}|\mu^i, \mu > \mu^i > \mu^*\right) &= \frac{\frac{1}{2}E\left(C_3^i|\mu^i, \mu > \mu^i > \mu^*\right) - M_1^*\left(\mu^i\right) - \frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2}{\frac{1}{2}I^i_{1,split}} \\ &= \frac{b\left(\mu^i\right)/\gamma - M_1^*\left(\mu^i\right)}{\frac{1}{2}I^i_{1,split}} \end{aligned}$$

with

$$b\left(\mu^i\right) = \gamma\frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1 + I^i_{1,split}\right) - I^i_{1,split}\right] - \frac{1}{8}\gamma^2\sigma^2\left(I^i_{1,split}\right)^2$$

and

$$1 + I^i_{1,split} = \frac{E\left(A_3|\mu^i\right)}{1 + \gamma\frac{1}{2}\sigma^2 I^i_{1,split}}.$$

We have already shown that $M_1^*\left(\mu^i\right)$ is a fraction of $b\left(\mu^i\right)/\gamma$. This is less than 1 and decreasing in μ^i. Thus, $E\left(r^i_{follow-on,final,risk-adj}|\mu^i,\mu>\mu^i>\mu^*\right)$ is positive. A sufficient condition for $E\left(r^i_{follow-on,final,risk-adj}|\mu^i,\mu>\mu^i>\mu^*\right)$ to be increasing in μ^i is that $\frac{b(\mu^i)}{I^i_{1,split}}$ is increasing in μ^i.

$$\begin{aligned}
\frac{d}{d\mu^i}\left(\frac{b\left(\mu^i\right)}{I^i_{1,split}}\right) &= \frac{\partial\left(\frac{b(\mu^i)}{I^i_{1,split}}\right)}{\partial I_1\left(\mu^i\right)}\frac{dI^i_{1,split}}{d\mu^i}+\frac{\partial\left(\frac{b(\mu^i)}{I^i_{1,split}}\right)}{\partial E\left(A_3|\mu^i\right)}\frac{dE\left(A_3|\mu^i\right)}{d\mu^i}\\
&= \frac{1}{\left(I^i_{1,split}\right)^2}\left(\frac{db\left(\mu^i\right)}{dI^i_{1,split}}I^i_{1,split}-b\left(\mu^i\right)\right)\frac{dI^i_{1,split}}{d\mu^i}\\
&\quad +\gamma\frac{1}{2}\frac{\ln\left(1+I^i_{1,split}\right)}{I^i_{1,split}}\frac{dE\left(A_3|\mu^i\right)}{d\mu^i}\\
&= \frac{1}{\left(I^i_{1,split}\right)^2}\left(-b\left(\mu^i\right)\right)\frac{dI^i_{1,split}}{d\mu^i}+\gamma\frac{1}{2}\frac{\ln\left(1+I^i_{1,split}\right)}{I^i_{1,split}}\frac{dE\left(A_3|\mu^i\right)}{d\mu^i}
\end{aligned}$$

where $\frac{db(\mu^i)}{dI^i_{1,split}}=0$ since $I^i_{1,split}$ maximizes $b\left(\mu^i\right)$. Furthermore, from the expression for $I^i_{1,split}$,

$$\frac{dI^i_{1,split}}{d\mu^i}=\frac{\frac{E\left(A_3|\mu^i\right)}{d\mu^i}}{\left[1+\gamma\frac{1}{2}\sigma^2+\gamma\sigma^2\left(I^i_{1,split}\right)^2\right]}$$

and

$$1+\gamma\frac{1}{2}\sigma^2\left(I^i_{1,split}\right)^2=E\left(A_3|\mu^i\right)-I^i_{1,split}\left(1+\gamma\frac{1}{2}\sigma^2\right).$$

Therefore,

$$
\begin{aligned}
0 &< \frac{d}{d\mu^i}\left(\frac{b\left(\mu^i\right)}{I^i_{1,split}}\right) \\
0 &< \frac{1}{\left(I^i_{1,split}\right)^2}\left(-b\left(\mu^i\right)\right)\frac{1}{\left[1+\gamma\frac{1}{2}\sigma^2+\gamma\sigma^2\left(I^i_{1,split}\right)\right]} \\
&\quad +\gamma\frac{1}{2}\frac{\ln\left(1+I^i_{1,split}\right)}{I^i_{1,split}} \\
b\left(\mu^i\right) &< \gamma\frac{1}{2}\ln\left(1+I^i_{1,split}\right)\left[I^i_{1,split}\left(1+\gamma\frac{1}{2}\sigma^2\right)+\gamma\sigma^2\left(I^i_{1,split}\right)^2\right] \\
0 &> \gamma\frac{1}{2}\left[\begin{array}{c}\left(E\left(A_3|\mu^i\right)-I^i_{1,split}\left(1+\gamma\frac{1}{2}\sigma^2\right)-\gamma\sigma^2\left(I^i_{1,split}\right)^2\right) \\ \ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\end{array}\right] \\
&\quad -\frac{1}{8}\gamma^2\sigma^2\left(I^i_{1,split}\right)^2
\end{aligned}
$$

$$
\begin{aligned}
0 &> \gamma\frac{1}{2}\left[\begin{array}{c}\left(1+\gamma\frac{1}{2}\sigma^2\left(I^i_{1,split}\right)^2-\gamma\sigma^2\left(I^i_{1,split}\right)^2\right) \\ \ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\end{array}\right] \\
&\quad -\frac{1}{8}\gamma^2\sigma^2\left(I^i_{1,split}\right)^2 \\
I^i_{1,split} &> \left(1-\gamma\frac{1}{2}\sigma^2\left(I^i_{1,split}\right)^2\right)\ln\left(1+I^i_{1,split}\right)-\gamma\frac{1}{4}\sigma^2\left(I^i_{1,split}\right)^2
\end{aligned}
$$

which is true since $I^i_{1,split} > \ln\left(1+I^i_{1,split}\right)$ for any $I^i_{1,split} > 0$ and $1-\gamma\frac{1}{2}\sigma^2\left(I^i_{1,split}\right)^2 < 1$ and $\gamma\frac{1}{4}\sigma^2\left(I^i_{1,split}\right)^2 > 0$.

Step 2: We then write $E\left(r^i_{follow-on,final,risk-adj}|r^i_{first,interim},\mu > \mu^i > \mu^*\right)$ as a function of $E\left(r^i_{follow-on,final,risk-adj}|\mu^i,\mu > \mu^i > \mu^*\right)$ and the distribution of μ^i conditional on $r^i_{first,interim}$.

$$
\begin{aligned}
& E\left(r^i_{follow-on,final,risk-adj}|r^i_{first,interim},\mu > \mu^i > \mu^*\right) \\
=& \int_{\mu^*}^{\mu} E\left(r^i_{follow-on,final,risk-adj}|\mu^i\right) f\left(\mu^i|r^i_{first,interim},\mu > \mu^i > \mu^*\right) d\mu^i.
\end{aligned}
$$

Note that

$$
1+r^i_{first,interim} = \frac{\frac{1}{2}E(C^i_2|H^i_1)-M_0}{\frac{1}{2}I_0} = \frac{\frac{1}{2}\left(a+H^i_1+E\left(H^i_2|H^i_1\right)\right)-M_0}{\frac{1}{2}I_0} = \frac{\frac{1}{2}\left(a+2H^i_1\right)-M_0}{\frac{1}{2}I_0}
$$

since $H_2^i = \mu^i + v^i = H_1^i - \varepsilon^i + v^i$. Therefore, $E\left(r^i_{follow-on,final,risk-adj} | r^i_{first,interim}, \mu^i > \mu^*\right)$ will be increasing in $r^i_{first,interim}$ iff $E\left(r^i_{follow-on,final,risk-adj} | H_1^i, \mu^i > \mu^*\right)$ is increasing in H_1^i. So we are interested in

$$
\begin{aligned}
& E\left(r^i_{follow-on,final,risk-adj} | H_1^i, \mu > \mu^i > \mu^*\right) \\
= & \int_{\mu^*}^{\mu} E\left(r^i_{follow-on,final,risk-adj} | \mu^i\right) f\left(\mu^i | H_1^i, \mu > \mu^i > \mu^*\right) d\mu^i
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{d}{dH_1^i} E\left(r^i_{follow-on,final,risk-adj} | H_1^i, \mu > \mu^i > \mu^*\right) \\
= & \int_{\mu^*}^{\mu} E\left(r^i_{follow-on,final,risk-adj} | \mu^i\right) \frac{df\left(\mu^i | H_1^i, \mu > \mu^i > \mu^*\right)}{dH_1^i} d\mu^i
\end{aligned}
$$

Since $H_1^i = \mu^i + \varepsilon^i$ with $\varepsilon^i \sim N\left(0, \sigma_\varepsilon^2\right)$, $\sigma_\varepsilon^2 = \frac{\frac{1}{2}\sigma^2\left(I_0^i\right)^2}{\left[\ln\left(1+I_0^i\right)\right]^2}$, we have $H_1^i | \mu^i, \mu > \mu^i > \mu^* \sim N\left(\mu^i, \sigma_\varepsilon^2\right)$ and

$$
\begin{aligned}
f\left(H_1^i | \mu > \mu^i > \mu^*\right) &= \int_{\mu^*}^{\mu} f\left(H_1^i | \mu^i, \mu > \mu^i > \mu^*\right) f\left(\mu^i | \mu > \mu^i > \mu^*\right) d\mu^i \\
&= \int_{\mu^*}^{\mu} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} d\mu^i \frac{1}{\mu - \mu^*} = \frac{1}{\mu - \mu^*}\left[\Phi\left(z_\mu\right) - \Phi\left(z_{\mu*}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
f\left(\mu^i | H_1^i, \mu > \mu^i > \mu^*\right) &= f\left(H_1^i | \mu^i, \mu > \mu^i > \mu^*\right) \frac{f\left(\mu^i | \mu > \mu^i > \mu^*\right)}{f\left(H_1^i | \mu > \mu^i > \mu^*\right)} \\
&= \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} \frac{\frac{1}{\mu-\mu^*}}{\frac{1}{\mu-\mu^*}\left[\Phi\left(z_\mu\right) - \Phi\left(z_{\mu*}\right)\right]} = \frac{\frac{1}{\sigma_\varepsilon}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right) - \Phi\left(z_{\mu*}\right)\right]}
\end{aligned}
$$

for $\mu > \mu^i > \mu^*$, 0 otherwise, where ϕ and Φ are the pdf and cdf of the standard normal distribution, $z_{\mu^i} = \frac{H_1^i - \mu^i}{\sigma_\varepsilon}$, $z_\mu = \frac{H_1^i - \mu}{\sigma_\varepsilon}$, and $z_{\mu^*} = \frac{H_1^i - \mu^*}{\sigma_\varepsilon}$. Note that this simply says that $\mu^i | H_1^i, \mu > \mu^i > \mu^*$ is truncated normal, with truncation at $-\mu$ and μ^*. Since $\phi\left(z_{\mu^i}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}$, $\frac{d\phi\left(z_{\mu^i}\right)}{dH_1^i} =$

$-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}\frac{z_{\mu^i}}{\sigma_\varepsilon} = -\phi\left(z_{\mu^i}\right)\frac{z_{\mu^i}}{\sigma_\varepsilon}$, we have

$$\begin{aligned}
\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i} &= \frac{-\frac{1}{\sigma_\varepsilon}\phi\left(z_{\mu^i}\right)\frac{z_{\mu^i}}{\sigma_\varepsilon}}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)} - \frac{\frac{1}{\sigma_\varepsilon}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right]^2}\left[\phi\left(z_\mu\right)\left(\frac{1}{\sigma_\varepsilon}\right)-\phi\left(z_{\mu^*}\right)\left(\frac{1}{\sigma_\varepsilon}\right)\right] \\
&= \frac{\frac{1}{\sigma_\varepsilon}\phi\left(z_{\mu^i}\right)\frac{1}{\sigma_\varepsilon}}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)}\left\{-z_{\mu^i}-\frac{\phi\left(z_\mu\right)-\phi\left(z_{\mu^*}\right)}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)}\right\} \\
&= f\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)\frac{1}{\sigma_\varepsilon}\left\{-z_{\mu^i}-\frac{\phi\left(z_\mu\right)-\phi\left(z_{\mu^*}\right)}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)}\right\}.
\end{aligned}$$

The function $f\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)\frac{1}{\sigma_\varepsilon}$ is positive for all values of μ^i. $\left\{-z_{\mu^i}-\frac{\phi(z_\mu)-\phi\left(z_{\mu^*}\right)}{\Phi(z_\mu)-\Phi(z_{\mu*})}\right\}$ is increasing in μ^i (as $\frac{\phi(z_\mu)-\phi\left(z_{\mu^*}\right)}{\Phi(z_\mu)-\Phi(z_{\mu*})}$ does not depend on i). Thus, there exists a value of μ^i, call it μ^x, which will depend on H_1^i and for which $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}=0$ for $\mu^i=\mu^x$, $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}<0$ for $\mu^i<\mu^x$, and $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}>0$ for $\mu^i>\mu^x$. It follows that $\frac{d}{dH_1^i}E\left(r^i_{follow-on,final,risk-adj}|H_1^i,\mu>\mu^i>\mu^*\right)=\int_{\mu^*}^{\mu}E\left(r^i_{follow-on,final,risk-adj}|\mu^i\right)\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}d\mu^i$ is positive (for all values of H_1^i) because (i) $\int_{\mu^*}^{\mu}\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}d\mu^i=0$ and (ii) $E\left(r^i_{follow-on,final,risk-adj}|\mu^i\right)$ is positive and increasing, implying that the positive values of $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}$ in the expression $\int_{\mu^*}^{\mu}E\left(r^i_{follow-on,final,risk-adj}|\mu^i\right)\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}d\mu^i$ are multiplied by a larger positive number than are the negative values of $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}$.

Proof of Implication 2

We prove Implication 2a for both our asymmetric information setup and for the symmetric information case in which both incumbent LPs and outside investors obtain the same information about the GP's type at $t=1$ (namely the hard information H_1^i). Implication 2b applies only in the asymmetric-learning setup of our model.

(a) Asymmetric-Learning Case: Note that

$$\begin{aligned}
1+r^i_{first,interim} &= \frac{\frac{1}{2}E(C_2^i|H_1^i)-M_0}{\frac{1}{2}I_0} = \frac{\frac{1}{2}\left(a+H_1^i+E\left(H_2^i|H_1^i\right)\right)\ln\left(1+I_0\right)-M_0}{\frac{1}{2}I_0} \\
&= \frac{\frac{1}{2}\left(a+2H_1^i\right)\ln\left(1+I_0\right)-M_0}{\frac{1}{2}I_0}.
\end{aligned}$$

This implies that $r^i_{first,interim}$ fully reveals H^i_1:

$$H^i_1 = \frac{1}{2}\left(\left[\left(1 + r^i_{first,interim}\right)\frac{1}{2}I_0 + M_0\right]\frac{1}{\frac{1}{2}\ln\left(1+I_0\right)} + a\right)$$

Since H^i_1 is positively related to $r^i_{first,interim}$, it follows that $P\left(\mu^i > \mu^*|r^i_{first,interim}, \mu > \mu^i > -\mu\right)$ is increasing in $r^i_{first,interim}$ iff $P\left(\mu^i > \mu^*|H^i_1, \mu > \mu^i > -\mu\right)$ is increasing in H^i_1.

Following steps similar to those in Step 2 of the proof of Implication 1,

$$f\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right) = \frac{\frac{1}{\sigma_\varepsilon}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)\right]}, \text{ for } \mu > \mu^i > -\mu, \ 0 \text{ otherwise,}$$

where $z_{\mu^i} = \frac{H^i_1 - \mu^i}{\sigma_\varepsilon}$, $z_\mu = \frac{H^i_1 - \mu}{\sigma_\varepsilon}$, and $z_{-\mu} = \frac{H^i_1 + \mu}{\sigma_\varepsilon}$ and

$$\frac{df\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right)}{dH^i_1} = f\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right)\frac{1}{\sigma_\varepsilon}\left\{z_{\mu^i} + \frac{\phi\left(z_\mu\right) - \phi\left(z_{-\mu}\right)}{\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)}\right\}.$$

We are interested in $P\left(\mu^i > \mu^*|H^i_1, \mu > \mu^i > -\mu\right) = \int_{\mu^*}^{\mu} f\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right)d\mu^i$ and in particular

$$\frac{dP\left(\mu^i > \mu^*|H^i_1, \mu > \mu^i > -\mu\right)}{dH^i_1} = \int_{\mu^*}^{\mu}\frac{df\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right)}{dH^i_1}d\mu^i.$$

The expression $f\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right)$ is positive for all values of μ^i between $-\mu$ and μ. The expression $\left\{z_{\mu^i} + \frac{\phi(z_\mu)-\phi(z_{-\mu})}{\Phi(z_\mu)-\Phi(z_{-\mu})}\right\}$ is increasing in μ^i (since $\frac{\phi(z_\mu)-\phi(z_{-\mu})}{\Phi(z_\mu)-\Phi(z_{-\mu})}$ does not depend on μ^i). Thus, there exists a value of μ^i, call it μ^x, which will depend on H^i_1 and for which $\frac{df\left(\mu^i|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1} = 0$ for $\mu^i = \mu^x$, $\frac{df\left(\mu^i|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1} < 0$ for $\mu^i < \mu^x$, and $\frac{df\left(\mu^i|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1} > 0$ for $\mu^i > \mu^x$. Since $\int_{-\mu}^{\mu}\frac{df\left(\mu^i|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1}d\mu^i = 0$, it follows that if $\mu^* > \mu^x$ then all values of $\frac{df\left(\mu^i|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1}$ in the integral for $\frac{dP\left(\mu^i>\mu^*|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1}$ are positive, so $\frac{dP\left(\mu^i>\mu^*|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1}$ is positive. If $\mu^* < \mu^x$, then $\frac{df\left(\mu^i|H^i_1,\mu>\mu^i>-\mu\right)}{dH^i_1} < 0$ for all $\mu^i < \mu^*$, so

$$\begin{aligned}\frac{dP\left(\mu^i > \mu^*|H^i_1, \mu > \mu^i > -\mu\right)}{dH^i_1} &= \int_{\mu^*}^{\mu}\frac{df\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right)}{dH^i_1}d\mu^i \\ &= 0 - \int_{-\mu}^{\mu^*}\frac{df\left(\mu^i|H^i_1, \mu > \mu^i > -\mu\right)}{dH^i_1}d\mu^i > 0.\end{aligned}$$

Symmetric-Information Case: To proceed with the proof for the symmetric information case, we must first state the solution of the model for that case. With symmetric learning, GPs have no reason to limit the number of LPs in a given fund. We assume that there is a mass of one of investors who each invests in all VC funds raised. With a continuum of GPs, all risk being idiosyncratic, and each LP investing in each of a continuum of GP types, all risk diversifies away from the perspective of a given LP. Each LP's wealth at $t = 3$ is:

$$\begin{aligned} W_3^{LP} &= W_0^{LP} + E_{\mu^i}\left[A_2^i \ln\left(1+I_0\right) - I_0 - M_0\right] + E_{H_1^i}\left[E\left(A_3^i \ln\left(1+I_1\right) - I_1 - M_1 | H_1^i\right)\right] \\ &= W_0^{LP} + \left[a \ln\left(1+I_0\right) - I_0 - M_0\right] + E_{H_1^i}\left[E\left(A_3^i \ln\left(1+I_1\right) - I_1 - M_1 | H_1^i\right)\right] \end{aligned}$$

Without informational hold-up, the market for funding remains competitive at all times. At $t = 1$, for a GP releasing hard information H_1^i concerning his first fund, the LPs' participation constraint for follow-on fund-raising is:

$$E\left(A_3^i | H_1^i\right) \ln\left(1 + I_1\left(H_1^i\right)\right) - I_1\left(H_1^i\right)\right) - M_1\left(H_1^i\right) = 0.$$

A given GP thus sets $M_1\left(H_1^i\right) = E\left(A_3^i | H_1^i\right) \ln\left(1 + I_1\left(H_1^i\right)\right) - I_1\left(H_1^i\right)\right)$. The GP then picks fund size to maximize $M_1\left(H_1^i\right)$, which is simply the NPV of the fund. Our informational structure implies that

$$E\left(A_3^i | H_1^i\right) = E\left(a + H_{2F}^i + H_{3F}^i | H_1^i\right) = E\left(a + 2\mu^i + \varepsilon_F^i + v_F^i | H_1^i\right) = a + 2E\left(\mu^i | H_1^i\right) = a + 2H_1^i.$$

Maximizing the NPV of the fund thus results in:

$$\begin{aligned} I_1\left(H_1^i\right)^{\text{Sym info}} &= E\left(A_3^i | H_1^i\right) - 1 = a + 2H_1^i - 1, \text{ for } H_1^i > \frac{1-a}{2}, \text{ zero otherwise.} \\ M_1\left(H_1^i\right)^{\text{Sym info}} &= \left(a + 2H_1^i - 1\right)\left(\ln\left(a + 2H_1^i\right) - 1\right) \text{ for } H_1^i > \frac{1-a}{2}, \text{ zero otherwise.} \end{aligned}$$

The outcome for first funds is similar. At $t = 0$, LPs' participation constraint is $a \ln\left(1 + I_0\right) - I_0 - M_0 = 0$. A given GP thus sets $M_0 = a \ln\left(1 + I_0\right) - I_0$. The GP then picks fund size to maximize this expression, which is simply the average NPV of the fund, averaging across possible GP types, resulting in

$$I_0^{\text{Sym info}} = a - 1, \; M_0^{\text{Sym info}} = a\left(\ln\left(a\right) - 1\right).$$

Thus,

$$1+r^{i}_{first,interim}=\frac{E(C^{i}_{2}|H^{i}_{1})-M^{\text{Sym info}}_{0}}{I^{\text{Sym info}}_{0}}=\frac{\left(a+2H^{i}_{1}\right)\ln(a)-a\left(\ln(a)-1\right)}{a-1}.$$

This implies that $r^{i}_{first,interim}$ fully reveals H^{i}_{1}. Since follow-on funds are raised iff $H^{i}_{1}>\frac{1-a}{2}$, this means that they are raised iff

$$r^{i}_{first,interim}>\frac{\ln(a)-a\left(\ln(a)-1\right)}{a-1}-1.$$

The right-hand-side expression is a constant that is known at $t=0$. Denote it by $r^{i,*}_{first,interim}$. Thus, $P\left(H^{i}_{1}>\frac{1-a}{2}|r^{i}_{first,interim}\right)=0$ for $r^{i}_{first,interim}\leq r^{i,*}_{first,interim}$ and $P\left(H^{i}_{1}>\frac{1-a}{2}|r^{i}_{first,interim}\right)=1$ for $r^{i}_{first,interim}>r^{i,*}_{first,interim}$, implying that the probability that a GP raises a follow-on fund is (weakly) increasing in the LP return of the GP's first fund, $r^{i}_{first,interim}$.

(b) Start from our assumptions that

$$\begin{aligned}C^{i}_{2}&=A^{i}_{2}\ln\left(1+I^{i}_{0}\right)\\A^{i}_{2}&=a+H^{i}_{1}+H^{i}_{2}=a+2\mu^{i}+\varepsilon^{i}+v^{i}\\H^{i}_{1}&=\mu^{i}+\varepsilon^{i},\ H^{i}_{2}=\mu^{i}+v^{i}.\end{aligned}$$

Note that

$$\begin{aligned}1+r^{i}_{first,final}&=\frac{C^{i}_{2}-2M_{0}}{I_{0}}=\frac{A^{i}_{2}\ln(1+I_{0})-2M_{0}}{I_{0}}=\frac{\left(a+H^{i}_{1}+H^{i}_{2}\right)\ln(1+I_{0})-2M_{0}}{I_{0}}\\1+r^{i}_{first,interim}&=\frac{E(C^{i}_{2}|H^{i}_{1})-2M_{0}}{I_{0}}=\frac{\left(a+H^{i}_{1}+E\left(H^{i}_{2}|H^{i}_{1}\right)\right)\ln(1+I_{0})-2M_{0}}{I_{0}}\\&=\frac{\left(a+2H^{i}_{1}\right)\ln(1+I_{0})-2M_{0}}{I_{0}}.\end{aligned}$$

This implies that $r^{i}_{first,interim}$ fully reveals H^{i}_{1} and given H^{i}_{1}, $r^{i}_{first,final}$ fully reveals H^{i}_{2}:

$$\begin{aligned}H^{i}_{1}&=\frac{1}{2}\left(\left[\left(1+r^{i}_{first,interim}\right)\frac{1}{2}I_{0}+M_{0}\right]\frac{1}{\frac{1}{2}\ln(1+I_{0})}+a\right)\\H^{i}_{2}&=\left(\left[\left(1+r^{i}_{first,final}\right)\frac{1}{2}I_{0}+M_{0}\right]\frac{1}{\frac{1}{2}\ln(1+I_{0})}+a+H^{i}_{1}\right)\end{aligned}$$

Since H_2^i is positively related to $r^i_{first,final}$, it follows that $P\left(\mu^i > \mu^* | r^i_{first,interim}, r^i_{first,final}, \mu > \mu^i > -\mu\right)$ is increasing in $r^i_{first,final}$ iff $P\left(\mu^i > \mu^* | H_1^i, H_2^i, \mu > \mu^i > -\mu\right)$ is increasing in H_2^i.

The information about μ^i in H_1^i and H_2^i can be summarized by the average $H^i = \frac{1}{2}\left(H_1^i + H_2^i\right) = \mu^i + \frac{1}{2}\left(\varepsilon^i + v^i\right)$. Thus, $\mu^i = H^i - \frac{1}{2}\left(\varepsilon^i + v^i\right)$ and $\mu^i | H_1^i, H_2^i \sim \mu^i | H^i$. Furthermore, $V\left(H^i | \mu^i\right)$ (denote it by σ_H^2) equals $\frac{1}{4}\left(\sigma_\varepsilon^2 + \sigma_v^2\right)$ and

$$\begin{aligned}
f\left(H^i | \mu > \mu^i > -\mu\right) &= \int_{\mu^*}^{\mu} f\left(H^i | \mu^i, \mu > \mu^i > -\mu\right) f\left(\mu^i | \mu > \mu^i > -\mu\right) d\mu^i \\
&= \int_{\mu^*}^{\mu} \frac{1}{\sqrt{2\pi\sigma_H^2}} e^{-\frac{1}{2}\left(z^{avg}_{\mu^i}\right)^2} d\mu^i \frac{1}{2\mu} = \frac{1}{2\mu}\left[\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)\right]
\end{aligned}$$

with $z^{avg}_{\mu^i} = \frac{H^i - \mu^i}{\sigma_H}$, $z^{avg}_{\mu} = \frac{H^i - \mu}{\sigma_H}$, and $z^{avg}_{-\mu} = \frac{H^i + \mu}{\sigma_H}$ (using "avg" to refer to these variables depending on the average value of H_1^i and H_2^i). Thus,

$$\begin{aligned}
f\left(\mu^i | H^i, \mu > \mu^i > -\mu\right) &= f\left(H^i | \mu^i, \mu > \mu^i > -\mu\right) \frac{f\left(\mu^i | \mu > \mu^i > -\mu\right)}{f\left(H^i | \mu > \mu^i > -\mu\right)} \\
&= \frac{1}{\sqrt{2\pi\sigma_H^2}} e^{-\frac{1}{2}\left(z^{avg}_{\mu^i}\right)^2} \frac{\frac{1}{2\mu}}{\frac{1}{2\mu}\left[\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)\right]} = \frac{\frac{1}{\sigma_H}\phi\left(z^{avg}_{\mu^i}\right)}{\left[\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)\right]}
\end{aligned}$$

for $\mu > \mu^i > -\mu$, and 0 otherwise. Thus, $\mu^i | H_1^i, H_2^i, \mu > \mu^i > -\mu$ has a truncated normal distribution with truncation at $-\mu$ and μ. Note that $\phi\left(z^{avg}_{\mu^i}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z^{avg}_{\mu^i}\right)^2}$ implies $\frac{d\phi\left(z^{avg}_{\mu^i}\right)}{dH_2^i} = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z^{avg}_{\mu^i}\right)^2} \frac{z^{avg}_{\mu^i}}{2\sigma_H} = -\phi\left(z^{avg}_{\mu^i}\right) \frac{z^{avg}_{\mu^i}}{2\sigma_H}$. Therefore,

$$\begin{aligned}
\frac{df\left(\mu^i | H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} &= \frac{-\frac{1}{\sigma_H}\phi\left(z^{avg}_{\mu^i}\right)\frac{z^{avg}_{\mu^i}}{2\sigma_H}}{\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)} \\
&\quad - \frac{\frac{1}{\sigma_H}\phi\left(z^{avg}_{\mu^i}\right)}{\left[\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)\right]^2}\left[\phi\left(z^{avg}_{\mu}\right)\left(\frac{1}{2\sigma_H}\right) - \phi\left(z^{avg}_{-\mu}\right)\left(\frac{1}{2\sigma_H}\right)\right] \\
&= \frac{\frac{1}{\sigma_H}\phi\left(z^{avg}_{\mu^i}\right)\frac{1}{2\sigma_H}}{\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)}\left\{-z^{avg}_{\mu^i} - \frac{\phi\left(z^{avg}_{\mu}\right) - \phi\left(z^{avg}_{-\mu}\right)}{\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)}\right\} \\
&= f\left(\mu^i | H_1^i, H_2^i, \mu > \mu^i > -\mu\right)\frac{1}{2\sigma_H}\left\{-z^{avg}_{\mu^i} - \frac{\phi\left(z^{avg}_{\mu}\right) - \phi\left(z^{avg}_{-\mu}\right)}{\Phi\left(z^{avg}_{\mu}\right) - \Phi\left(z^{avg}_{-\mu}\right)}\right\}.
\end{aligned}$$

We are interested in $P\left(\mu^i > \mu^* | H_1^i, H_2^i, \mu > \mu^i > -\mu\right) = \int_{\mu^*}^{\mu} f\left(\mu^i | H_1^i, H_2^i, \mu > \mu^i > -\mu\right) d\mu^i$ and in particular,

$$\frac{dP\left(\mu^i > \mu^*|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} = \int_{\mu^*}^{\mu} \frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} d\mu^i.$$

The expression $f\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)$ is positive for all values of μ^i between $-\mu$ and μ. The expression $\left\{-z_{\mu^i}^{avg} - \frac{\phi\left(z_{\mu}^{avg}\right)-\phi\left(z_{-\mu}^{avg}\right)}{\Phi\left(z_{\mu}^{avg}\right)-\Phi\left(z_{-\mu}^{avg}\right)}\right\}$ is increasing in μ^i (since $\frac{\phi\left(z_{\mu}^{avg}\right)-\phi\left(z_{-\mu}^{avg}\right)}{\Phi\left(z_{\mu}^{avg}\right)-\Phi\left(z_{-\mu}^{avg}\right)}$ does not depend on μ^i). Thus, there exists a value of μ^i, call it μ^x, which will depend on H^i and for which $\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i} = 0$ for $\mu^i = \mu^x$, $\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i} < 0$ for $\mu^i < \mu^x$, and $\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i} > 0$ for $\mu^i > \mu^x$. Since $\int_{-\mu}^{\mu} \frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i} d\mu^i = 0$, it follows that if $\mu^* > \mu^x$ then all values of $\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i}$ in the integral for $\frac{dP\left(\mu^i>\mu^*|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i}$ are positive, so $\frac{dP\left(\mu^i>\mu^*|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i}$ is positive. If $\mu^* < \mu^x$, then $\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>-\mu\right)}{dH_2^i} < 0$ for all $\mu^i < \mu^*$, so

$$\begin{aligned} \frac{dP\left(\mu^i > \mu^*|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} &= \int_{\mu^*}^{\mu} \frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} d\mu^i \\ &= 0 - \int_{-\mu}^{\mu^*} \frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} d\mu^i > 0. \end{aligned}$$

Proof of Implication 3

As was the case for Implication 2a, we prove Implication 3a for both our asymmetric-learning setup and the symmetric-information version of our model in which both incumbent and outside investors obtain the same information about the GP's type at $t = 1$ (namely, the hard information H_1^i). Implication 3b applies only in the asymmetric-learning case.

(a) Asymmetric-Learning Case: If raised, the follow-on fund's size is

$$I_{1,split}^i = \frac{-\left(1+\gamma\frac{1}{2}\sigma^2\right) + \sqrt{\left(1+\gamma\frac{1}{2}\sigma^2\right)^2 - 2\gamma\sigma^2\left[1 - E\left(A_3^i|\mu^i\right)\right]}}{\gamma\sigma^2}$$

where $E\left(A_2^i|\mu^i\right) > 1$ for $\mu^i > \mu^*$. Since $A_2^i = a + H_1^i + H_2^i = a + 2\mu^i + \varepsilon^i + v^i$, we have $E\left(A_2^i|\mu^i\right) = a + 2\mu^i$, so $I_{1,split}^i$ is positive and increasing in μ^i.

Since H_1^i is positively related to $r^i_{first,interim}$,

$$H_1^i = \frac{1}{2}\left(\left[\left(1+r^i_{first,interim}\right)\frac{1}{2}I_0 + M_0\right]\frac{1}{\frac{1}{2}\ln\left(1+I_0\right)} + a\right)$$

it follows that $E\left(I^i_{1,split}|r^i_{first,interim}, \mu > \mu^i > -\mu^*\right)$ is increasing in $r^i_{first,interim}$ iff $E\left(I^i_{1,split}|H^i_1, \mu > \mu^i > -\mu^*\right)$ is increasing in H^i_1. Furthermore,

$$\begin{aligned}
E\left(I^i_{1,split}|H^i_1, \mu > \mu^i > -\mu^*\right) &= \int_{\mu^*}^{\mu} I^i_{1,split} f\left(\mu^i|H^i_1, \mu > \mu^i > \mu^*\right) d\mu^i \\
\frac{dE\left(I^i_{1,split}|H^i_1, \mu > \mu^i > -\mu^*\right)}{dH^i_1} &= \int_{\mu^*}^{\mu} I^i_{1,split} \frac{df\left(\mu^i|H^i_1, \mu > \mu^i > \mu^*\right)}{dH^i_1} d\mu^i
\end{aligned}$$

This is positive following the same arguments as in Step 2 of the proof of Implication 1 (simply replace $E\left(r^i_{follow-on,final,risk-adj}|\mu^i\right)$ by $I^i_{1,split}$).

Symmetric-Information Case: From the proof of Implication 2a for the symmetric-information case, we have that if raised, i.e., if $H_1^i > \frac{1-a}{2}$, the follow-on fund's size is

$$I_1\left(H_1^i\right)^{\text{Sym info}} = a + 2H_1^i - 1$$

and

$$1 + r^i_{first,interim} = \frac{\left(a+2H_1^i\right)\ln\left(a\right) - a\left(\ln\left(a\right)-1\right)}{a-1}.$$

Combining these two expressions, we get

$$I_1\left(H_1^i\right)^{\text{Sym info}} = \frac{\left(1+r^i_{first,interim}\right)\left(a-1\right) + a\left(\ln\left(a\right)-1\right)}{\ln\left(a\right)} - 1$$

which is a linear and increasing function of $r^i_{first,interim}$. Since $E\left(I_1^i|r^i_{first,interim}, H_1^i > \frac{1-a}{2}\right) = I_1\left(H_1^i\right)^{\text{Sym info}}$, this proves the implication for the symmetric-information case.

(b) If raised, the follow-on fund's size is

$$I^i_{1,split} = \frac{-\left(1+\gamma\frac{1}{2}\sigma^2\right) + \sqrt{\left(1+\gamma\frac{1}{2}\sigma^2\right)^2 - 2\gamma\sigma^2\left[1 - E\left(A_2^i|\mu^i\right)\right]}}{\gamma\sigma^2}$$

where $E\left(A_2^i|\mu^i\right) > 1$ for $\mu^i > \mu^*$. Since $A_2^i = a + H_1^i + H_2^i = a + 2\mu^i + \varepsilon^i + v^i$, we have $E\left(A_2^i|\mu^i\right) = a + 2\mu^i$. Thus, $I_{1,split}^i$ is positive and increasing in μ^i.

From the expressions in the above proof expressing H_1^i, H_2^i as linear increasing functions of $r_{first,interim}^i$ and $r_{first,final}^i$, it follows that $E\left(I_{1,split}^i|r_{first,interim}^i, r_{first,final}^i, \mu > \mu^i > -\mu^*\right)$ is increasing in $r_{first,final}^i$ iff $E\left(I_{1,split}^i|H_1^i, H_2^i, \mu > \mu^i > -\mu^*\right)$ is increasing in H_2^i. Furthermore,

$$
\begin{aligned}
E\left(I_{1,split}^i|H_1^i, H_2^i, \mu > \mu^i > -\mu^*\right) &= \int_{\mu^*}^{\mu} I_{1,split}^i f\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right) d\mu^i \\
\frac{dE\left(I_{1,split}^i|H_1^i, H_2^i, \mu > \mu^i > -\mu^*\right)}{dH_2^i} &= \int_{\mu^*}^{\mu} I_{1,split}^i \frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right)}{dH_2^i} d\mu^i.
\end{aligned}
$$

From the same steps as in the above proof,

$$
f\left(\mu^i|H^i, \mu > \mu^i > \mu^*\right) = \frac{\frac{1}{\sigma_H}\phi\left(z_{\mu^i}^{avg}\right)}{\left[\Phi\left(z_{\mu}^{avg}\right) - \Phi\left(z_{\mu^*}^{avg}\right)\right]}, \text{ for } \mu > \mu^i > \mu^*, \text{ 0 otherwise.}
$$

with $z_{\mu^i}^{avg} = \frac{H^i - \mu^i}{\sigma_H}$, $z_{\mu}^{avg} = \frac{H^i - \mu}{\sigma_H}$, and $z_{\mu}^{avg} = \frac{H^i + \mu^*}{\sigma_H}$. Thus, again using the same steps as in the above proof,

$$
\frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} = f\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right)\frac{1}{2\sigma_H}\left\{-z_{\mu^i}^{avg} - \frac{\phi\left(z_{\mu}^{avg}\right) - \phi\left(z_{\mu^*}^{avg}\right)}{\Phi\left(z_{\mu}^{avg}\right) - \Phi\left(z_{\mu^*}^{avg}\right)}\right\}.
$$

The expression $f\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right)$ is positive for all values of μ^i. $\left\{-z_{\mu^i}^{avg} - \frac{\phi\left(z_{\mu}^{avg}\right) - \phi\left(z_{\mu^*}^{avg}\right)}{\Phi\left(z_{\mu}^{avg}\right) - \Phi\left(z_{\mu^*}^{avg}\right)}\right\}$ is increasing in μ^i (since $\frac{\phi\left(z_{\mu}^{avg}\right) - \phi\left(z_{\mu^*}^{avg}\right)}{\Phi\left(z_{\mu}^{avg}\right) - \Phi\left(z_{\mu^*}^{avg}\right)}$ does not depend on i). Thus, there exists a value of μ^i, call it μ^x (which will depend on H^i) for which $f\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right) = 0$ for $\mu^i = \mu^x$, $f\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right) < 0$ for $\mu^i < \mu^x$, and $f\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right) > 0$ for $\mu^i > \mu^x$. Therefore, $\frac{dE\left(I_{1,split}^i|H_1^i, H_2^i, \mu > \mu^i > -\mu^*\right)}{dH_2^i}$ is positive (for all values of H_0^i) since $\int_{\mu^*}^{\mu} \frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right)}{dH_0^i} d\mu^i = 0$ and $I_{1,split}^i$ is positive and increasing, implying that the positive values of $\frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right)}{dH_2^i}$ in the expression $\int_{\mu^*}^{\mu} I_{1,split}^i \frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right)}{dH_2^i} d\mu^i$ are multiplied by a larger positive number than are the negative values of $\frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > \mu^*\right)}{dH_2^i}$.

Proof of Implication 4

The final return on a follow-on fund is

$$1 + r^i_{follow-on,final} = \frac{C^i_3 - 2M^*_1\left(\mu^i\right)}{I^i_{1,split}}.$$

From the proof of Implication 1 we know that $E\left(r^i_{follow-on,final}|\mu^i, \mu > \mu^i > \mu^*\right)$ is positive and increasing in μ^i, since we showed that $E\left(r^i_{follow-on,final,risk-adj}|\mu^i, \mu > \mu^i > \mu^*\right)$ is increasing in μ^i and since the risk adjustment $\gamma\frac{1}{4}\sigma^2\left(I^i_{1,split}\right)^2$ is positive and increasing in μ^i. From the expressions in the proof of Implication 2 expressing H^i_1, H^i_2 as linear increasing functions of $r^i_{first,interim}$ and $r^i_{first,final}$, it follows that $E\left(r^i_{follow-on,final,risk-adj}|r^i_{first,interim}, r^i_{first,final}, \mu > \mu^i > -\mu^*\right)$ is increasing in $r^i_{first,final}$ iff $E\left(r^i_{follow-on,final,risk-adj}|H^i_1, H^i_2, \mu > \mu^i > -\mu^*\right)$ is increasing in H^i_2. Furthermore,

$$\begin{aligned}
& E\left(r^i_{follow-on,final,risk-adj}|H^i_1, H^i_2, \mu > \mu^i > -\mu^*\right) \\
= \quad & E\left(r^i_{follow-on,final,risk-adj}|\mu^i, \mu > \mu^i > \mu^*\right) f\left(\mu^i|H^i_1, H^i_2, \mu > \mu^i > \mu^*\right) d\mu^i
\end{aligned}$$

$$\begin{aligned}
& \frac{dE\left(r^i_{follow-on,final,risk-adj}|H^i_1, H^i_2, \mu > \mu^i > -\mu^*\right)}{dH^i_2} \\
= \quad & \int_{\mu^*}^{\mu} E\left(r^i_{follow-on,final,risk-adj}|\mu^i, \mu > \mu^i > \mu^*\right) \frac{df\left(\mu^i|H^i_1, H^i_2, \mu > \mu^i > \mu^*\right)}{dH^i_2} d\mu^i
\end{aligned}$$

This is positive following the same steps as in the proof of Implication 3.

Figure 1. Fund Risk

The figure shows the distribution of net IRRs for mutual funds, hedge funds, VC funds, and buyout funds in the U.S. for the period from 1980 to 2006. The graphs present, for each set of funds, Gaussian kernel densities of net annual IRRs from CRSP (for mutual funds), the CISDM Hedge Funds database available on WRDS (for hedge funds), PREQIN (for buyout funds), and combination of PREQIN and Venture Economics (for VC funds). The unit of observation in the hedge-fund and mutual fund kernels is a fund-year; the unit of observation in the other two kernels is a fund, as VC and buyout funds last 10 years. The data contain 48,314 observations for hedge funds, 222,205 observations for mutual funds, 1,208 observations for VC funds, and 669 observations for buyout funds. Net IRRs in excess of 200% p.a. exist but are not shown.

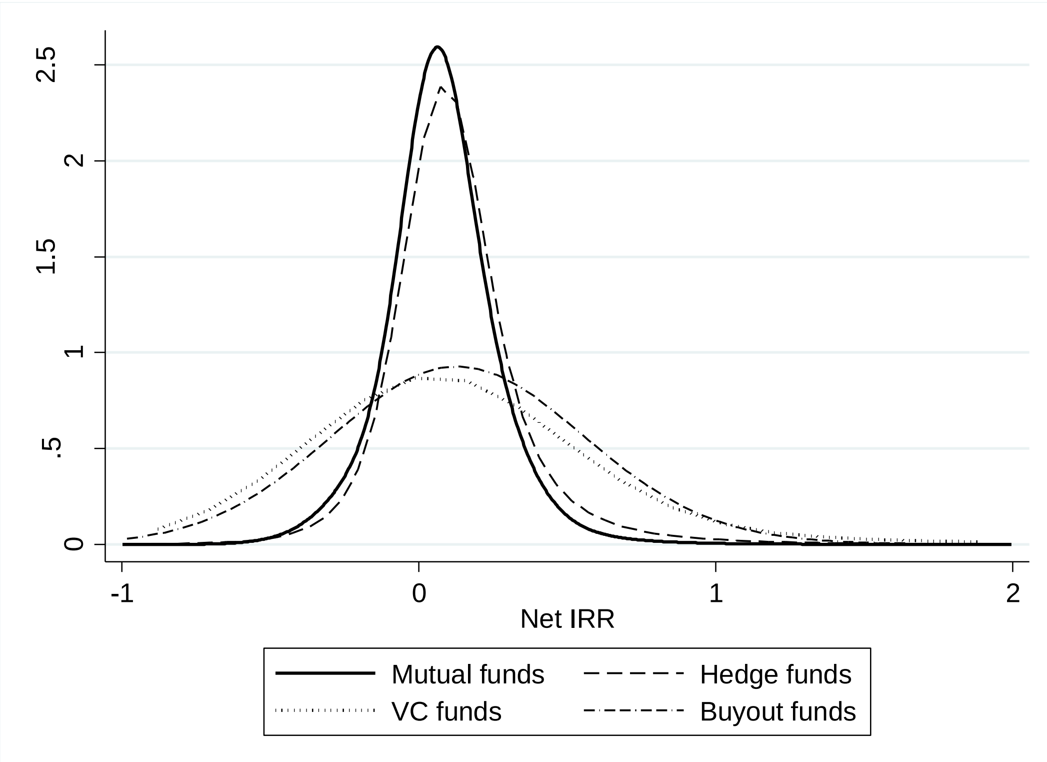
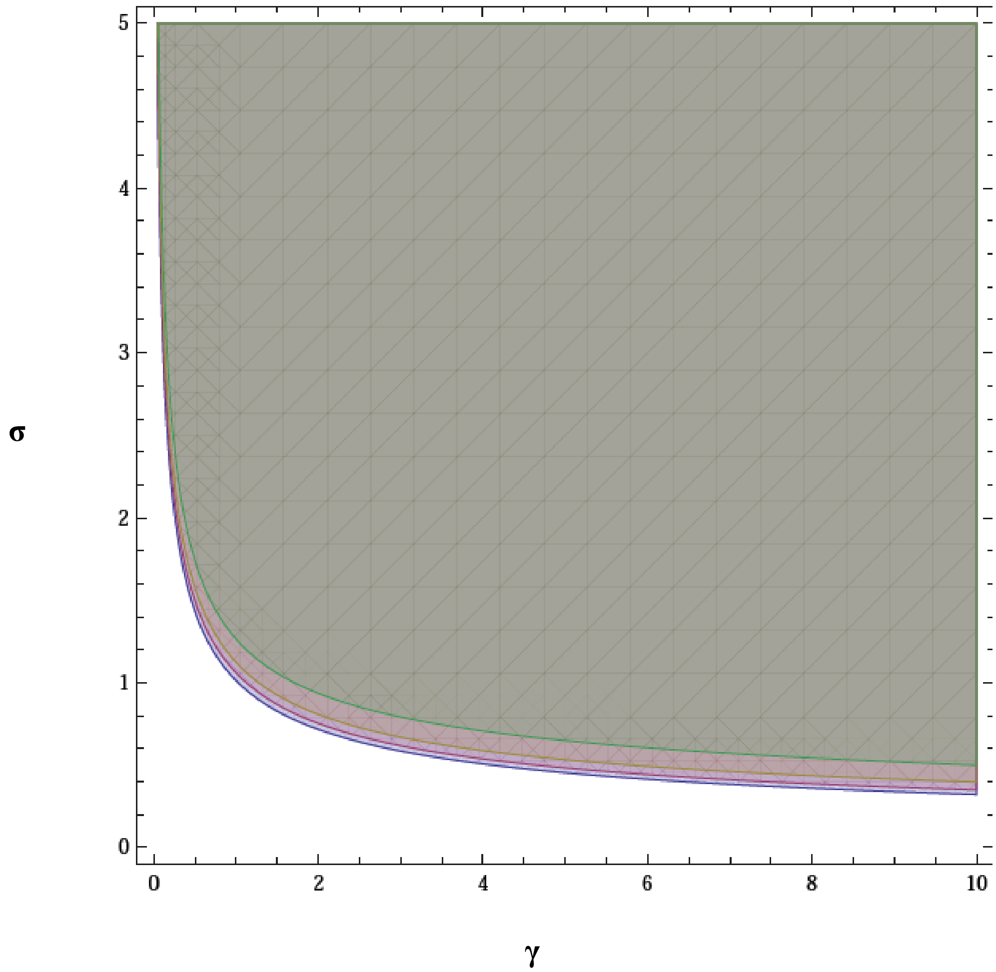


Figure 2. Illustration of Corollary 1.

The figure illustrates Corollary 1 by depicting the values of γ and σ (shaded area) for which the condition in Proposition 1 holds, for values of $E(A_3|\mu^i)$ in the set [1.1 1.5 2 3].



$E(A_3|\mu^i) = 1.1$; $E(A_3|\mu^i) = 1.5$; $E(A_3|\mu^i) = 2$; $E(A_3|\mu^i) = 3$.

Figure 3. Average Interim IRRs Over a Fund's Lifetime.
The figure shows the average interim IRR, net of fees, in percent over a fund's 10-year lifetime for a sample of 547 VC funds for which a complete time series of year-by-year interim IRR data is available.

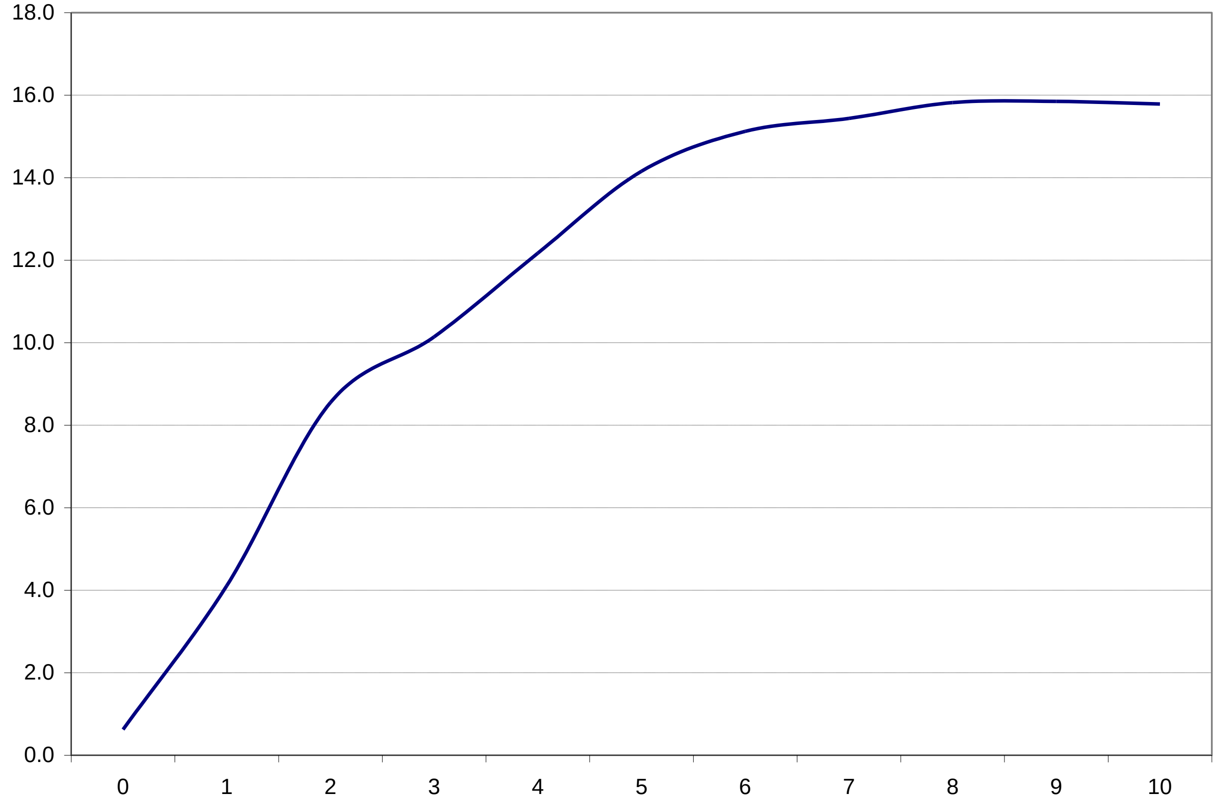


Figure 4. How Accurately Do Interim IRRs Forecast Final Performance?

The figure shows box plots of the distribution of forecast errors (= final IRR – interim IRR, in %) for each year in a fund's life, using all 15,205 fund-years for which interim IRR are available. Each box shows the 75th percentile (the upper hinge of the box), the median (the line drawn inside the box), and the 25th percentile (the lower hinge).The whiskers extending from each box denote the 5th and 95th percentiles.

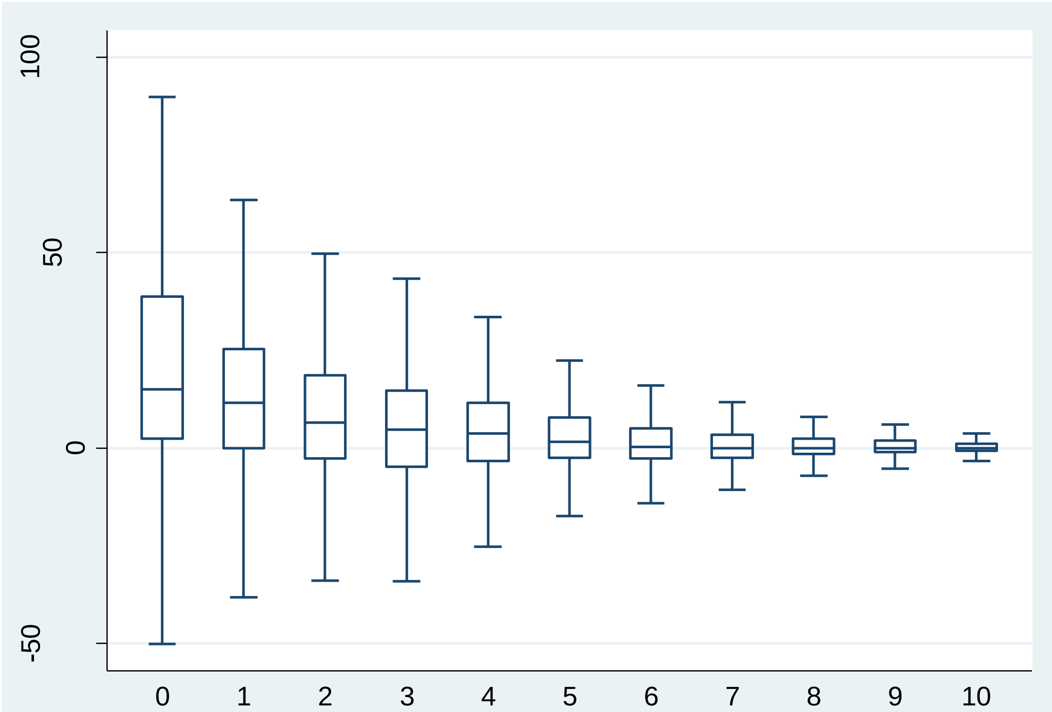


Figure 5. Early Versus Late-Stage Funds

The figure shows the distribution of final IRRs for early- and late-stage VC funds in our sample, respectively. Each graph presents a Gaussian kernel density using optimal half-widths and 100 estimation points.

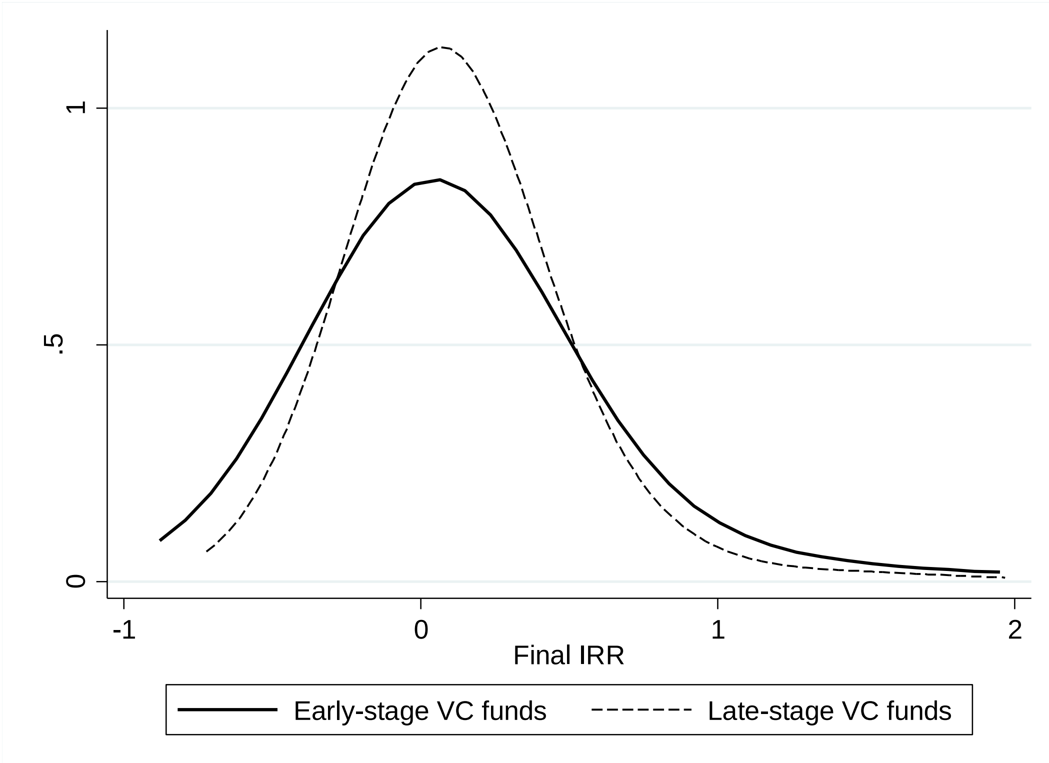


Table 1. Survey Evidence: Do LPs Receive Priority in Follow-on Funds and If So, Why?

Da Rin and Phalippou (2011) conduct a survey of 2,000 limited partners in private equity and venture capital funds between 2008 and 2010. The response rate is in excess of 10%. Survey question 3.7 is directly relevant to our model, and this table reproduces the answers. Results look nearly identical if only the responses of U.S.-based LPs are tabulated.

	Percent of LPs who answered:					
Question:	Always	Sometimes	Never	Do not know	Yes (Always+Sometimes)	N
3.7 In your experience, does investing in a fund give you priority over other investors when the GP raises subsequent funds?	44.4	43.1	7.5	5.0	87.5	239

	Percent of LPs who answered:					
Question:	No	Yes, possibly	Yes, definitely	Do not know	Yes (Yes, possibly + Yes, definitely)	N
If yes, why do you think you receive priority?						
3.7.1 If I didn't re-invest, other investors would be suspicious and would not invest.	17.4	56.7	15.4	10.5	72.1	201
3.7.2 If the GP didn't allow me to reinvest, I could replicate their strategy (myself or in cooperation with another GP).	80.3	11.1	2.0	6.6	13.1	198

Table 2. Descriptive Statistics.

The sample consists of 2,257 U.S. venture capital funds raised by 962 VC firms between 1980 and 2002, as reported Venture Economics (VE) and Private Equity Intelligence (PREQIN). We define as VC funds all funds listed in VE or PREQIN as focusing on start-up, seed, early-stage, development, late-stage, or expansion investments, as well as those listed as "venture (general)" or "balanced" funds. In cases where VE and PREQIN classify a fund differently, we verify fund type using secondary sources such as *Pratt's Guide*, *CapitalIQ*, *Galante's*, and a web search. We screen out funds of funds, buyout funds, hedge funds, venture leasing funds, evergreen funds (i.e., funds without a predetermined dissolution date), and side funds. Fund size is in nominal dollars. A first-time fund is the first fund raised by a VC firm, assigned fund sequence number 1. Subsequent follow-on funds are numbered accordingly. Early-stage funds are those focused on start-up, seed, early-stage, or development investments. The final IRR is a fund's annual internal rate of return estimated over its (typically 10-year) life, net of management and performance fees, using VE and PREQIN data through October 2012.

	Number of sample funds	of which			Fund size ($m)					Performance			
vintage	all	only in VE	only in PREQIN	in both	mean	median	fraction first-time funds	mean fund sequence no.	fraction early-stage funds	no. of funds with final IRR data	mean final IRR (%)	sd final IRR (%)	median final IRR (%)
1980	37	31	4	2	30.4	20.0	0.68	1.4	0.35	17	13.0	12.7	12.9
1981	46	38	1	7	25.4	20.0	0.70	1.6	0.35	20	11.1	15.5	10.4
1982	62	51	1	10	24.8	15.6	0.71	1.5	0.37	29	5.2	14.4	6.5
1983	71	58	1	12	33.2	21.0	0.46	1.8	0.41	42	8.6	11.4	7.8
1984	81	67	1	13	33.9	23.4	0.47	2.0	0.43	54	5.2	9.6	3.9
1985	58	39	1	18	41.2	20.0	0.38	2.1	0.47	32	10.6	10.9	12.1
1986	55	36	1	18	54.6	22.0	0.47	2.1	0.49	34	8.5	8.1	6.6
1987	78	62	0	16	35.7	23.6	0.41	2.2	0.40	55	7.0	15.1	7.2
1988	56	31	2	23	67.9	32.8	0.27	2.5	0.54	41	15.2	15.5	12.5
1989	75	42	1	32	68.0	30.5	0.36	2.7	0.51	51	16.3	31.8	12.2
1990	45	33	2	10	46.0	35.0	0.40	2.8	0.49	19	17.0	21.6	13.7
1991	32	21	1	10	43.4	35.0	0.31	2.4	0.47	16	23.6	18.0	22.6
1992	50	28	0	22	79.0	49.1	0.22	3.2	0.42	29	22.7	27.5	13.2
1993	73	43	3	27	56.2	35.9	0.34	2.7	0.41	40	27.6	32.7	19.3
1994	72	38	0	34	86.1	46.5	0.26	3.0	0.50	42	23.3	32.9	17.0
1995	113	77	1	35	72.3	50.0	0.42	2.6	0.56	55	44.0	58.0	27.2
1996	95	64	1	30	71.6	50.0	0.47	2.6	0.52	43	59.3	99.3	20.8
1997	162	100	3	59	84.0	57.0	0.44	2.8	0.51	75	40.6	72.3	9.5
1998	171	103	1	67	137.3	74.5	0.28	3.3	0.58	83	25.8	100.7	3.9
1999	249	163	2	84	171.9	100.0	0.36	3.2	0.65	82	-5.1	13.9	-5.2
2000	332	213	2	117	201.4	100.0	0.35	3.2	0.65	110	-2.1	12.8	-1.7
2001	171	110	4	57	209.6	61.5	0.35	3.2	0.64	57	-1.7	10.8	-0.6
2002	73	43	4	26	130.2	45.0	0.32	3.6	0.52	26	-3.4	9.3	-2.5
1980-2002	2,257	1,491	37	729	111.2	46.0	0.39	2.8	0.54	1,052	15.7	47.6	5.6

Table 3. VC Fund Performance Persistence.

This table reports tests of Implications 1 and 4 of the model, regarding performance persistence across funds managed by the same VC firm. We regress the ex post performance of fund *N* on the performance of the fund manager's previous fund (*N*-1) and controls for fund size (in log $m) and risk (an indicator for funds with a focus on early-stage ventures). The dependent variable in columns 1 and 3 through 6 is a fund's ex post IRR, net of carry and fees, measured at the end of the fund's usually ten-year life. (The sample accordingly consists of funds that are at least ten years old as of 2012, that is, funds raised between 1980 and 2002.) In column 2, we measure performance using exit rates, defined as the fraction of a fund's investments that were exited through an IPO or an M&A transaction over the course of the fund's ten-year life. The performance of a fund manager's previous fund is measured either ex post (i.e., after ten years) or using the "interim" IRR that the previous fund reported in the year before fund *N* was raised. In terms of the model, ex post returns are considered "soft" information and interim returns are considered "hard" information. Columns 1 and 2 replicate Kaplan and Schoar's results (2005) using IRRs and exit rates, respectively. Columns 3 and 4 test Implication 1a and 1b, respectively. Columns 5 and 6 test Implication 4. All models are estimated using OLS with vintage-year fixed effects. Heteroskedasticity-consistent standard errors, clustered on VC firm, are shown in italics. We use ***, **, and * to denote significance at the 1%, 5%, and 10% level (two-sided), respectively.

	Ex post performance of fund *N*					
	IRR	Exit rate	IRR			
Performance measure:	(1)	(2)	(3)	(4)	(5)	(6)
Previous fund's performance						
ex post IRR or exit rate of fund *N*-1	0.247***	0.319***			0.302***	0.301***
	0.068	*0.036*			*0.068*	*0.069*
interim IRR of fund *N*-1 as of previous year			0.110***	0.104***	0.060**	0.058**
			0.038	*0.036*	*0.028*	*0.028*
Controls						
log size of fund *N*-1	0.055***	0.019***	0.069***	0.077***	0.061***	0.059***
	0.014	*0.006*	*0.023*	*0.023*	*0.022*	*0.021*
dummy = 1 if fund *N* has early-stage focus				0.088*	0.069*	0.069*
				0.049	*0.041*	*0.041*
years since raising fund *N*-1						-0.004
						0.016
Diagnostics						
Vintage year FE	yes	yes	yes	yes	yes	yes
Wald test: all coeff. = 0	7.5***	10.9***	7.3***	6.4***	8.1***	8.2***
Adjusted R^2	16.3%	17.2%	16.2%	16.7%	23.0%	22.8%
No. of observations	628	1,079	387	387	374	374

Table 4. Effect of Learning on Fund-raising.

This table reports tests of Implications 2 and 3 of the model, regarding the effect of performance on future fund-raising. In columns 1 and 2, we estimate a Cox semi-parametric hazard model with time-varying covariates using annual data. This models the hazard (i.e., the instantaneous probability) that a VC firm raises a new fund in year *t*. We allow a VC firm to raise multiple funds in succession by estimating a "multiple-failure" hazard model. Column 1 conditions on the size and interim IRR of the VC firm's "current" fund with meaningful returns, both as of the end of year *t*-1. (The current fund is the VC firm's highest-numbered fund that is at least 3 years old and in operation at *t*-1.) Thus, this hazard model uses only information that was publicly available to incumbent LP and outside investors at the time of fund-raising. It includes all available vintages through 2012; since VC firms have a non-zero probability of raising further funds after that date, the hazard model adjusts for right-censoring. Column 2 adds soft information available to incumbent LPs (but not outside investors) in the form of the ex post IRR on the VC firm's current fund as of year *t*-1. This is a measure of soft information about the GP's performance. Columns 3 and 4 estimate the size of a follow-on fund. The dependent variable is the log of the size of the follow-on fund (in $m) if the firm raises a follow-on fund and zero if it does not. To code failure to raise a follow-on fund, we identify 661 defunct VC firms in *CapitalIQ*. The model is estimated using Tobit. Column 3 focuses on the interim IRR of the previous fund measured as of the year-end prior to the year the GP raises the current fund; if no follow-on fund is raised, the IRR of the previous fund is measured ex post (i.e., as of year ten.) Column 4 adds the previous fund's ex post IRR. Standard errors are shown in italics. They are clustered on VC firm in columns 1 and 2; the Tobit estimator in columns 3 and 4 does not support clustering. We use ***, **, and * to denote significance at the 1%, 5%, and 10% level (two-sided), respectively.

	Prob(follow-on fund raised)	*Prob*(follow-on fund raised)	Log size of follow-on fund	Log size of follow-on fund
	(1)	(2)	(3)	(4)
Previous fund's performance				
interim IRR of fund *N-1* as of previous year-end	0.270***	0.132***	2.178***	1.785***
	0.040	*0.059*	*0.260*	*0.294*
ex post IRR of fund *N*-1		0.226***		0.450**
		0.067		*0.221*
Controls				
log fund size	0.195***	0.187***	1.386***	1.417***
	0.025	*0.025*	*0.098*	*0.101*
Diagnostics				
Vintage year FE	n.a.	n.a.	yes	yes
Wald test: all coeff. = 0	125.7***	127.1***	359.4***	288.1***
Pseudo R^2	n.a.	n.a.	10.8%	9.7%
No. of observations	3,880	3,874	767	684
No. of VC firms	302	301		
No. of funds raised	771	770		
Model estimated	Hazard	Hazard	Tobit	Tobit

Table 5. Alternative Explanation for Persistence: Fundraising in Good and Bad Times.

This table tests an informal alternative explanation for performance persistence: GPs give incumbent LPs a share of the rents to ensure stable relationships over time, so that fundraising is easier in bad times. Under this explanation, performance persistence should disappear if one focuses on VC firms that have raised funds in both "bad" and "good" fundraising years. Unlike in Table 3, the sample is therefore restricted to VC firms that have raised funds in both "bad" and "good" years over the sample period. Column 1 defines "bad" years as those in which total fundraising in the U.S. VC industry declined by at least 10% in dollar terms compared to the year before (i.e., 1985, 1987, 1990, 1991, 1996, 2001, and 2002). Column 2 defines "bad" years as those in which fundraising in the VC industry declined by at least 20% compared to the year before (i.e., 1990, 1991, 2001, and 2002). Column 3 defines "bad" years as those in which fewer first-time funds were raised than in the year before (i.e., 1983, 1985, 1988, 1990, 1991, 1994, 1996, 1998, 2001, and 2002). Column 4 defines "bad" years as those in which fewer follow-on funds were raised than in the year before (i.e., 1985, 1986, 1988, 1990, 1991, 1996, 2001, and 2002). "Good" years are those not classified as "bad". We regress fund *N*'s ex post IRR, net of carry and fees, measured at the end of the fund's ten-year life, on the performance of the fund manager's previous fund (*N*-1) and controls for fund size and risk. The performance of a fund manager's previous fund is measured either ex post (i.e., after ten years) or using the "interim" IRR that the previous fund reported in the year before fund *N* was raised. All models are estimated using OLS with vintage-year fixed effects. Heteroskedasticity-consistent standard errors, clustered on VC firm, are shown in italics. We use ***, **, and * to denote significance at the 1%, 5%, and 10% level (two-sided), respectively.

	Ex post IRR of fund *N*			
	(1)	(2)	(3)	(4)
Previous fund's performance				
ex post IRR of fund *N*-1	0.311***	0.381***	0.293***	0.323***
	0.074	*0.087*	*0.066*	*0.072*
interim IRR of fund *N*-1 as of previous year	0.054**	0.074**	0.073***	0.056**
	0.027	*0.030*	*0.028*	*0.028*
Controls				
log size of fund *N*-1	0.055**	0.003	0.051**	0.056***
	0.024	*0.023*	*0.023*	*0.023*
dummy = 1 if fund *N* has early-stage focus	0.096**	0.071	0.071	0.091*
	0.048	*0.065*	*0.045*	*0.048*
Diagnostics				
Vintage year FE	yes	yes	yes	yes
Wald test: all coeff. = 0	11.3***	16.1***	8.1***	9.3***
Adjusted R^2	22.7%	23.8%	23.9%	23.6%
No. of observations	302	201	344	308

Internet Appendix. A Simplified Model with up to Three Funds Raised by Each GP

Hochberg, Ljungqvist, and Vissing-Jorgensen's (2012) informational hold-up model assumes that each GP raises at most two funds. In practice, GPs often raise more than two funds over time. Would informational hold-up imply that persistence remains even when comparing returns on, say, funds 2 and 3?

To examine this, we construct a simplified version of the model (with risk-neutral agents and non-overlapping funds) in which we allow each GP to raise up to three funds. This simplified version demonstrates that performance persistence is present both from fund 1 to 2 and from fund 2 to 3. Intuitively, performance persistence extends to later funds because only a small amount of information asymmetry is required to induce outside investors to withdraw from the market. It does not matter whether the information asymmetry is reduced over time as the performance of later funds is observed. What matters is simply that the information asymmetry remains positive.

1 Setup

Fund structure and general partner skill: A given fund is raised and managed by a single GP. GPs differ in skill.

At $t = 0$, each GP raises a fund of size I_0, which we refer to as a first-time fund. Depending on the information learned between $t = 0$ and $t = 1$, the GP may raise a follow-on fund of size I_1 at $t = 1$ and another one of size I_2 at $t = 2$. Each fund lasts one period. We assume that third funds can be raised only if second funds are raised.[1]

A GP's skill determines the properties of the cash flows returned by his funds and is captured by the variable μ^i. For a given GP i, the cash flows are

$$C_1^i = A_1^i \ln\left(1 + I_0^i\right) \text{ for the first fund} \quad (1)$$

$$C_2^i = A_2^i \ln\left(1 + I_1^i\right) \text{ for the second fund, if raised} \quad (2)$$

$$C_3^i = A_3^i \ln\left(1 + I_2^i\right) \text{ for the third fund, if raised} \quad (3)$$

[1] This could be motivated, for example, by GP skill depreciating to zero if he is out of the VC market for a period.

with subscripts referring to when the cash flow is received and where

$$A_1^i = a + \mu^i + \varepsilon_1^i, \ \varepsilon_1^i | \mu^i \sim N\left(0, \frac{\sigma^2 \left(I_0^i\right)^2}{\left[\ln\left(1+I_0^i\right)\right]^2}\right) \tag{4}$$

$$A_2^i = a + \mu^i + \varepsilon_2^i, \ \varepsilon_2^i | \mu^i \sim N\left(0, \frac{\sigma^2 \left(I_1^i\right)^2}{\left[\ln\left(1+I_1^i\right)\right]^2}\right) \tag{5}$$

$$A_3^i = a + \mu^i + \varepsilon_3^i, \ \varepsilon_3^i | \mu^i \sim N\left(0, \frac{\sigma^2 \left(I_2^i\right)^2}{\left[\ln\left(1+I_2^i\right)\right]^2}\right) \tag{6}$$

$$\varepsilon_1^i, \ \varepsilon_2^i, \ \varepsilon_3^i \text{ are independent, given } \mu^i. \tag{7}$$

All shocks are drawn independently across GPs and are independent of the GP's skill μ^i. All risk is idiosyncratic.

There is a continuum of GP types of mass one. We assume that μ^i is distributed uniformly over the interval $[-\mu, \mu]$ such that $\mu^i = 0$ corresponds to average skill. We abstract from agency problems by assuming that GPs manage their funds in their LPs' best interest.

Limited partners: There is a large set of ex ante identical investors, such that the LP market is perfectly competitive at $t = 0$. Each fund has one LP.

Preferences and wealth: Both GPs and LPs are risk-neutral and consume at time $t = 3$. Each GP has initial wealth of W_0^{GP} and each LP has initial wealth of W_0^{LP}. In addition to investing in the VC industry, LPs can invest at a riskfree rate of r_f, set equal to zero without loss of generality. We assume that each LP can invest in one first-time fund and, if desired, in any follow-on fund raised by the same GP. Cash flows received at $t = 1$ and $t = 2$ are invested at the riskfree rate until $t = 3$.

Learning about GP skill: At time $t = 0$, no-one knows the GP's skill, μ^i. At $t = 1$, the GP and the LP who invested in the GP's first fund learn the GP's skill μ^i . LPs who have not invested in the GP's first fund only observe its cash flow (and fund size, I_0). We refer to this setup as asymmetric learning, in the sense that the incumbent LP learns the GP's type faster than do outside investors.

Payoff functions: We assume that the cash flow from a given fund is divided between the GP and the LP according to the following contract agreed at the start of the fund. For first-time funds, GP i receives (at the end of the fund, at $t = 1$) a dollar amount of

$$X_1^{GP} = M_1. \tag{8}$$

The LP investing with this GP receives, at $t = 1$,

$$X_1^{LP} = C_1^i - M_1. \tag{9}$$

Similar, for second or third funds, payoffs are:

$$X_2^{GP} = M_2, \ X_2^{LP} = C_2^i - M_2 \text{ at } t = 2 \tag{10}$$

$$X_3^{GP} = M_3, \ X_3^{LP} = C_3^i - M_3 \text{ at } t = 3. \tag{11}$$

While the fees in the model are expressed as dollar amounts, once the optimal fund size has been derived, one can calculate the implied percentage fee for a given fund, which then corresponds to the management fee used in actual VC contracts.

The values of the fees, the fund sizes, and which LPs invest in follow-on funds will be the focus of the solution of the model.

2 Fund Size and Fee in Follow-On Funds (Second or Third Funds)

Under asymmetric learning, the LP market is perfectly competitive at $t = 0$ but not at $t = 1$. Because outside investors do not learn the GP's type, the incumbent LP has an informational advantage over outside investors when the GP attempts to raise a follow-on fund. This allows the incumbent LP to extract part of the follow-on fund's value. While it is intuitive that the incumbent LP's informational advantage should improve the terms he obtains, it is useful to model the bargaining game explicitly since it allows us to be clear about the role played by outside investors.

Bargaining: We use a bargaining setup based on Rubinstein (1982) bargaining, adapted to the VC setting. For a particular GP i and his incumbent LP, we assume the following sequential bargaining game starting at $t = 1$ for second funds. The same game is then repeated at $t = 2$ for third funds (replace subscript 1 by 2 and subscript 2 by 3 for third funds).

(1) The GP makes an offer to the incumbent LP to invest I_1^{GP} and pay a fee of M_2^{GP}.

(2) If the GP's offer is rejected, the incumbent LP offers to provide funds I_1^{LP} and pay a fee of M_2^{LP}.

(3) If the LP's offer is rejected, the GP makes another offer; and so on.

We assume that delay in reaching an agreement is costly and, as is standard in Rubinstein bargaining (e.g., Binmore, Rubinstein, and Wolinsky (1986)), model this by assuming that between each round of offers there is an exogenous probability p that the bargaining process will terminate without an agreement.

If no agreement is reached, each party receives its outside option. For the incumbent LP, this equals a riskfree return of r_f. The GP's outside option depends on what outside investors are willing to offer if no agreement is reached with the incumbent LP. We assume that outside investors cannot see (or cannot verify) the bids made prior to breakdown of bargaining with the incumbent LP. Therefore, they do not know whether bargaining has broken down for exogenous reasons, or whether it has broken down because one of the parties has simply refused to bargain any further.

We furthermore assume that an incumbent LP can always counter any offer an outside investor makes. The GP's outside option is then zero, because outside investors face a winner's curse. Say an outside investor observes a first-time fund cash flow of C_1^i. Denote by μ^* the value of μ such that the NPV of the GP's fund is zero. The outside investor knows that if he offers an investment and fee that gives the LP an expected payoff above his outside option when the GP's type μ^i exceeds μ^*, the incumbent LP will counter with an offer that is more attractive to the GP only when μ^i in fact exceeds μ^*. When the GP's type μ^i is below μ^*, no LP offer can be made that yields a payoff to the incumbent LP above his outside option and that the GP will accept, because the fund NPV is negative (implying that the total surplus to be shared between the GP and LP is negative). Therefore, for $\mu^i < \mu^*$, the incumbent LP does not counter any outside LP offer, leaving the outside investor with a loss. Understanding this, the outside investor can never make an expected-utility-increasing investment and rationally withdraws from the market.

The fund sizes that maximizes joint surplus: To solve for the Nash equilibrium strategies, it is helpful to start by deriving the fund sizes for second and third funds that maximize the joint surplus of the GP and LP. With risk-neutral agents, this will simply be the fund size that maximizes the fund's NPV. The fund sizes solve

$$\max_{I_1, I_2} E\left(U^{GP}|\mu^i\right) + E\left(U^{LP}|\mu^i\right) \tag{12}$$

where

$$E\left(U^{GP}|\mu^i\right) = W_3^{GP} = W_0^{GP} + M_1 + M_2 + M_3 \tag{13}$$

$$\begin{aligned} E\left(U^{LP}|\mu^i\right) &= E\left(W_3^{LP}|\mu^i\right) = W_0^{LP} + \left(E\left(A_1^i|\mu^i\right)\ln\left(1+I_0\right) - M_1 - I_0\right) \\ &+ \left(E\left(A_2^i|\mu^i\right)\ln\left(1+I_1\right) - M_2 - I_1\right) + \left(E\left(A_3^i|\mu^i\right)\ln\left(1+I_2\right) - M_3 - I_2\right) \end{aligned} \tag{14}$$

Note that the fund sizes that maximize the joint surplus do not depend on the fees, M_2 and M_3. Instead, these fees simply determine how the surplus is shared. Maximizing the joint surplus thus implies solving:

$$\max_{I_1}\left(E\left(A_2|\mu^i\right)\ln\left(1+I_1\right) - I_1\right) \text{ and } \max_{I_2}\left(E\left(A_3|\mu^i\right)\ln\left(1+I_2\right) - I_2\right). \tag{15}$$

Since $E\left(A_2|\mu^i\right) = E\left(A_3|\mu^i\right) = a + \mu^i$, the solution is

$$I_1\left(\mu^i\right) = I_2\left(\mu^i\right) = a + \mu^i - 1 \ . \tag{16}$$

Denote this common fund size by $I\left(\mu^i\right)$. As would be expected, the fund size that maximizes the joint surplus increases in GP skill. It equals zero for $\mu^i = 1 - a$. Therefore μ^*, the value for which the NPV of the fund is zero, is $1 - a$.

The LP's expected cash flow before fees in a second fund is

$$\left[E\left(A_2|\mu^i\right)\ln\left(1 + I\left(\mu^i\right)\right) - I\left(\mu^i\right)\right] \tag{17}$$

which is zero for $E\left(A_2|\mu^i\right) = 1$ (i.e., for $\mu^i = \mu^*$) and positive for $E\left(A_2|\mu^i\right) > 1$ ($\mu^i > \mu^*$). The solution for third funds is similar.

Nash equilibrium strategies, fund size, and fee: The following proposition states the equilibrium outcome.

Proposition 1: As $p \longrightarrow 0$, the following is a subgame perfect equilibrium in the bargaining game for second funds and in the bargaining game for third funds, with $j = 2$ for second funds and $j = 3$ for third funds:

(a) All offers involve fund sizes that maximize the joint surplus, $I\left(\mu^i\right)$ (which is zero for $\mu^i < \mu^*$).

(b) Denote by $M_j^{LP,*}$ and $M_j^{GP,*}$ the fees such that (i) the LP is indifferent between accepting the GP's offer and having his own offer accepted in the next round, and (ii) the GP is indifferent between accepting the LP's offer and having his own offer accepted in the next round. As $p \longrightarrow 0$, $M_j^{LP,*}$ and $M_j^{GP,*}$ both converge to $M\left(\mu^i\right) = \frac{1}{2}\left[E\left(A_j|\mu^i\right)\ln\left(1 + I\left(\mu^i\right)\right) - I\left(\mu^i\right)\right] = \frac{1}{2}\left[\left(a + \mu^i\right)\ln\left(1 + I\left(\mu^i\right)\right) - I\left(\mu^i\right)\right]$.

(c) The GP's strategy is to always offer $(I\left(\mu^i\right), M_j^{GP,*})$ and always reject offers with a fee below $M_j^{LP,*}$. The LP's strategy is to always offer $(I\left(\mu^i\right), M_j^{LP,*})$ whenever it is his turn to make an offer and always reject offers with a fee above $M_j^{GP,*}$.

Given (c), the equilibrium outcome of the sequential bargaining game is immediate agreement with the LP accepting the GP's first offer. The fund fee for $p \longrightarrow 0$ is thus $M\left(\mu^i\right)$ and the fund size is $I\left(\mu^i\right)$.

Proof of Proposition 1: See Proof section below. We focus on the case where $p \longrightarrow 0$ from here on.

The implication of Proposition 1 is that the LP earns rents in the form of an expected return in excess of the riskfree rate (set to zero for simplicity). In dollar terms, this rent is the same for second and third funds and is given by $\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right)-I\left(\mu^i\right)\right]=\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]$, which is simply half of the NPV of running the fund. Thus, the GP and LP share the value of running the second and third fund equally.

These rents increase in the GP's skill μ^i since

$$\begin{aligned} & \frac{d}{d\mu^i}\left\{\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]\right\} \qquad (18) \\ = & \frac{1}{2}\left(\ln\left(a+\mu^i\right)\right)>0 \text{ for } a+\mu^i>1 \text{ which is true for } \mu^i>\mu^*. \end{aligned}$$

Similarly, the expected return to the LP, after fees, in both second and third funds, is:

$$E\left(r_2^i|\mu^i,\mu>\mu^i>\mu^*\right) = E\left(r_3^i|\mu^i,\mu>\mu^i>\mu^*\right)=\frac{\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right)-I\left(\mu^i\right)\right]}{I\left(\mu^i\right)} \qquad (19)$$

$$= \frac{\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]}{\left(a+\mu^i-1\right)}. \qquad (20)$$

This also increases in μ^i since

$$\begin{aligned} & \frac{d}{d\mu^i}\left\{\frac{\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]}{\left(a+\mu^i-1\right)}\right\} \qquad (21) \\ = & \frac{\frac{1}{2}\ln\left(a+\mu^i\right)\left(a+\mu^i-1\right)-\left[\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]\right]}{\left(a+\mu^i-1\right)^2} \end{aligned}$$

which is positive when $\ln\left(a+\mu^i\right)\left(a+\mu^i-1\right)-\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]=\left(a+\mu^i-1\right)-\ln\left(a+\mu^i\right)>0$ which is true since $x-1>\ln(x)$ for any $x>1$.

We show below that this is what generates performance persistence.

3 Fund Size and Fee in First-Time Funds

As no learning has taken place yet, the LP market is perfectly competitive at time $t=0$. Accordingly, LPs have no bargaining power and all GPs offer LPs a contract that ensures the highest possible expected utility for the GP, subject to each LP achieving an expected utility across investing in both the GP's first and follow-on funds equal to what the LP could obtain by not investing in venture capital. We refer to this as the LP's participation constraint. This constraint will depend on the outcome for follow-on funds stated in Proposition 1.

We first determine the LP's participation constraint and then solve for the fund size that maximizes the GP's expected utility subject to this constraint. Not surprisingly, the fund size that results will be the one that maximizes joint GP and LP surplus, and this will be the fund size that maximizes the fund's NPV (as was the case for follow-on funds).

LP's participation constraint: The LP's expected utility is $EU^{LP} = E\left(W_3^{LP}\right) = E_{\mu^i}\left(E\left(W_3^{LP}|\mu^i\right)\right)$ and the LP's participation constraint is that $EU^{LP} = W_0^{LP}$. As of $t = 0$, GP type is unknown and so I_0 will not depend on μ^i. However, when calculating the LP's expected utility, expectations must be taken both with respect to μ^i and with respect to the shocks A_1, A_2, and A_3 conditional on μ^i. Furthermore, for follow-on funds, only funds with GP skill $\mu^i \geq \mu^*$ are raised, and thus $\left(E\left(A_2^i|\mu^i\right)\ln\left(1+I_1\left(\mu^i\right)\right) - M_2\left(\mu^i\right) - I_1\left(\mu^i\right)\right)$ and $\left(E\left(A_3^i|\mu^i\right)\ln\left(1+I_2\left(\mu^i\right)\right) - M_3\left(\mu^i\right) - I_2\left(\mu^i\right)\right)$ are zero for $\mu^i < \mu^*$. Therefore, the constraint is:

$$E_{\mu^i}\begin{pmatrix} W_0^{LP} + \left(E\left(A_1^i|\mu^i\right)\ln\left(1+I_0\right) - M_1 - I_0\right) \\ +\left(E\left(A_2^i|\mu^i\right)\ln\left(1+I_1\left(\mu^i\right)\right) - M_2\left(\mu^i\right) - I_1\left(\mu^i\right)\right) \\ +\left(E\left(A_3^i|\mu^i\right)\ln\left(1+I_2\left(\mu^i\right)\right) - M_3\left(\mu^i\right) - I_2\left(\mu^i\right)\right) \end{pmatrix} = W_0^{LP} \tag{22}$$

$\iff$

$$a\ln\left(1+I_0\right) - M_1 - I_0 + 2\int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M\left(\mu^i\right) - I\left(\mu^i\right)\right)\frac{1}{2\mu}d\mu^i = 0 \tag{23}$$

$\iff$

$$\begin{aligned} M_1\left(I_0\right) &= a\ln\left(1+I_0\right) - I_0 + 2\int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M\left(\mu^i\right) - I\left(\mu^i\right)\right)\frac{1}{2\mu}d\mu^i && (24) \\ &= a\ln\left(1+I_0\right) - I_0 + \int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\ln\left(a+\mu^i\right) - \left(a+\mu^i-1\right)\right)\frac{1}{2\mu}d\mu^i && (25) \end{aligned}$$

The LP's participation constraint simply says that, to the extent that the LP earns an expected cash flow after fees in follow-on funds that exceeds his investment (due to his informational hold-up power), he will earn an expected cash flow after fees in first funds that is below his investment in these funds.

Fund size and fee in first-time funds: The GP picks I_0 to maximize his expected utility subject to the LP's participation constraint:

$$\max_{I_0} W_0^{GP} + M_1 + E_{\mu^i}\left(M_2\left(\mu^i\right) + M_3\left(\mu^i\right)\right) \text{ s.t. } M_1 = M_1\left(I_0\right). \tag{26}$$

Since $W_0^{GP} + E_{\mu^i}\left(M_2\left(\mu^i\right) + M_3\left(\mu^i\right)\right)$ does not depend on what happens in the first-time fund, this implies simply choosing the value of I_0 that maximizes $M_1\left(I_0\right)$. This will be the value of I_0 that maximizes the NPV of the fund:

$$\begin{aligned} &\max_{I_0} a\ln\left(1+I_0\right) - I_0 + 2\int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M\left(\mu^i\right) - I\left(\mu^i\right)\right)\frac{1}{2\mu}d\mu^i && (27) \\ \iff &\max_{I_0} a\ln\left(1+I_0\right) - I_0 && (28) \\ \iff &I_0 = a - 1. && (29) \end{aligned}$$

The first-time fund fee is then given by $M_1(I_0)$ for this value of I_0 :

$$M_1(I_0) = a\ln(a) - (a-1) + \int_{\mu^*}^{\mu} \left(\left(a+\mu^i\right)\ln\left(a+\mu^i\right) - \left(a+\mu^i-1\right)\right)\frac{1}{2\mu}d\mu^i. \tag{30}$$

4 Empirical Implications

The realized returns to LPs after fees are given by:

$$\begin{aligned} r_1^i &= \frac{A_1^i\ln(1+I_0) - M_1(I_0) - I_0}{I_0} = \frac{\left(\mu^i+\varepsilon_1^i\right)\ln(a) - \frac{1}{2\mu}\int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\left(\ln\left(a+\mu^i\right)-1\right)+1\right)d\mu^i}{a-1} \\ r_2^i &= \frac{A_2^i\ln\left(1+I\left(\mu^i\right)\right) - M\left(\mu^i\right) - I\left(\mu^i\right)}{I\left(\mu^i\right)} = \frac{\frac{1}{2}\left[\left(a+\mu^i+\varepsilon_2^i\right)\ln\left(a+\mu^i\right) - \left(a+\mu^i-1\right)\right]}{a+\mu^i-1} \\ r_3^i &= \frac{A_3^i\ln\left(1+I\left(\mu^i\right)\right) - M\left(\mu^i\right) - I\left(\mu^i\right)}{I\left(\mu^i\right)} = \frac{\frac{1}{2}\left[\left(a+\mu^i+\varepsilon_3^i\right)\ln\left(a+\mu^i\right) - \left(a+\mu^i-1\right)\right]}{a+\mu^i-1} \end{aligned}$$

We next show that this model implies the persistence in LP returns after fees that has been documented in the literature.

Implication 1: Persistence in LP after-fee returns

(a) In the cross-section of GPs with follow-on funds, a high return to the LP (after fees) in a GP's first fund predicts a high return to the LP (after fees) in the GP's second fund: $E\left(r_2^i|r_1^i, \mu > \mu^i > \mu^*\right)$ is increasing in r_1^i.

(b) In the cross-section of GPs with follow-on funds, a high return to the LP (after fees) in a GP's second fund predicts a high return to the LP (after fees) in the GP's third fund: $E\left(r_3^i|r_2^i, \mu > \mu^i > \mu^*\right)$ is increasing in r_2^i.

(c) Persistence is stronger from second to third funds than from first to second funds: The regression coefficient on r_2 in a linear regression of values of r_3 on values of r_2 (and a constant term) is larger than the regression coefficient on r_1 in a linear regression of values of r_2 on values of r_1 (and a constant term). This is the case for all parameter values σ^2, μ, and a (with $a > 1$).

Proof: See Proof section below.

One might think that outside investors could simply invest in all second or third funds raised by GPs with high returns on their prior funds, thus expecting to earn high expected returns. Our model, however, makes it clear why this is not possible. The winner's curse problem described earlier implies that outside investors would only be able to invest with those GPs for whom their offers implied negative NPV to investors. This implies that the 'return-chasing' behavior emphasized by Berk and Green (2004) as the mechanism eliminating performance persistence in mutual funds breaks down in the VC setting when there is asymmetric learning.

Proofs for the Simplified Model with up to Three Funds Raised by Each GP

Proof of Proposition 1

For third funds:

(a): This part of the proposition is true for any value of p. Consider an offer (I_2^{GP}, M_3^{GP}) with fund size I_2^{GP} different from $I\left(\mu^i\right)$. By definition of $I\left(\mu^i\right)$ as the fund size that maximizes the joint surplus, the GP can always make himself better off by changing the proposed fund size to $I\left(\mu^i\right)$ and adjusting the proposed fee to keep the LP happy. A similar argument applies to offers made by the LP.

(b): The fees $M_3^{LP,*}$ and $M_3^{GP,*}$ that make the GP and the LP indifferent between accepting the other party's offer now and having their own offer accepted in the next offer round solve the following two equations. For any p, the GP's indifference condition is

$$M_3^{LP,*} = (1-p)\, M_3^{GP,*}.$$

The LP's indifference condition is

$$\begin{aligned} & E\left(A_3^i|\mu^i\right) \ln\left(1 + I\left(\mu^i\right)\right) - M_3^{GP,*} - I\left(\mu^i\right) \\ = \;& (1-p)\left[E\left(A_3^i|\mu^i\right) \ln\left(1 + I\left(\mu^i\right)\right) - M_3^{LP,*} - I\left(\mu^i\right)\right]. \end{aligned}$$

Combining the two equations implies:

$$\begin{aligned} M_3^{GP,*} &= \frac{E\left(A_3^i|\mu^i\right) \ln\left(1 + I\left(\mu^i\right)\right) - I\left(\mu^i\right)}{2-p} \\ M_3^{LP*} &= (1-p)\, M_3^{GP,*} = (1-p)\frac{E\left(A_3^i|\mu^i\right) \ln\left(1 + I\left(\mu^i\right)\right) - I\left(\mu^i\right)}{2-p} \end{aligned}$$

Thus, as p goes to zero:

$$\begin{aligned} M_3^{GP,*} &\to \frac{E\left(A_3^i|\mu^i\right) \ln\left(1 + I\left(\mu^i\right)\right) - I\left(\mu^i\right)}{2} \\ M_3^{LP,*} &\to \frac{E\left(A_3^i|\mu^i\right) \ln\left(1 + I\left(\mu^i\right)\right) - I\left(\mu^i\right)}{2}. \end{aligned}$$

We denote this common value by $M_3\left(\mu^i\right)$.

(c) We need to show that each party's strategy is an optimal response to the other party's strategy. By construction of $M_3^{GP,*}$ and $M_3^{LP,*}$ (as stated above), neither the GP nor the LP can hope to do better by rejecting the other party's offer. Furthermore, the GP cannot do better by increasing his proposed fee above $M_3^{GP,*}$, as the LP's strategy rejects all such offers. Additionally, the LP cannot do better by decreasing his proposed fee below $M_3^{LP,*}$ as the GP's strategy rejects all such offers.

For second funds:

(a): Same argument as for third funds.

(b): At the time the second fund is raised, both the GP and LP know what the bargaining outcome will be for the third fund, as derived above: Immediate acceptance by the LP of the GP's first offer with fee $M_3^{GP,*}$. Furthermore, given our assumption that third funds can be raised only if second funds are raised, a breakdown in negotiations for a second fund implies a loss of any payoff from third funds.

Thus, the fees $M_2^{LP,*}$ and $M_2^{GP,*}$ that make the GP and the LP indifferent between accepting the other party's offer now or having their own offer accepted in the next offer round solve the following two equations. For any p, the GP's indifference condition is

$$\begin{aligned} M_2^{LP,*} + M_3^{GP,*} &= (1-p)\left(M_2^{GP,*} + M_3^{GP,*}\right) \Longleftrightarrow \\ M_2^{LP,*} &= (1-p)\,M_2^{GP,*} - pM_3^{GP,*}. \end{aligned}$$

The LP's indifference condition is

$$\begin{aligned} &\left[E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M_2^{GP,*} - I\left(\mu^i\right)\right] + \left[E\left(A_3^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M_3^{GP,*} - I\left(\mu^i\right)\right] \\ = \;& (1-p)\left(\left[E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M_2^{LP,*} - I\left(\mu^i\right)\right] + \left[E\left(A_3^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M_3^{GP,*} - I\left(\mu^i\right)\right]\right). \end{aligned}$$

Combining the two equations implies:

$$\begin{aligned} &\left[E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M_2^{GP,*} - I\left(\mu^i\right)\right] + \left[E\left(A_3^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M_3^{GP,*} - I\left(\mu^i\right)\right] \\ = \;& (1-p)\begin{pmatrix} \left[E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - \left[(1-p)\,M_2^{GP,*} - pM_3^{GP,*}\right] - I\left(\mu^i\right)\right] \\ + \left[E\left(A_3^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - M_3^{GP,*} - I\left(\mu^i\right)\right] \end{pmatrix} \end{aligned}$$

$\Longleftrightarrow$

$$\begin{aligned} &2\left[E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - I\left(\mu^i\right)\right] - M_2^{GP,*} - M_3^{GP,*} \\ = \;& (1-p)\left(2\left[E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - I\left(\mu^i\right)\right] - (1-p)\left[M_2^{GP,*} + M_3^{GP,*}\right]\right) \end{aligned}$$

$\Longleftrightarrow$

$$(2-p)\left[M_2^{GP,*} + M_3^{GP,*}\right] = 2\left[E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - I\left(\mu^i\right)\right]$$

$\Longleftrightarrow$

$$\begin{aligned} M_2^{GP,*} &= \frac{E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - I\left(\mu^i\right)}{2-p}, \text{ which equals } M_3^{GP,*} \\ M_2^{LP*} &= (1-2p)\,\frac{E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right) - I\left(\mu^i\right)}{2-p}. \end{aligned}$$

Thus, as p goes to zero:

$$\begin{aligned} M_2^{GP,*} &\rightarrow \frac{E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right)-I\left(\mu^i\right)}{2} \\ M_2^{LP,*} &\rightarrow \frac{E\left(A_2^i|\mu^i\right)\ln\left(1+I\left(\mu^i\right)\right)-I\left(\mu^i\right)}{2}. \end{aligned}$$

We denote this common value by $M_2\left(\mu^i\right)$ (which equals $M_3\left(\mu^i\right)$).

(c) Same argument as for third funds.

Proof of Implication 1a

$$\begin{aligned} r_1^i &= \frac{A_1^i\ln\left(1+I_0\right)-M_1\left(I_0\right)-I_0}{I_0}=\frac{\left(\mu^i+\varepsilon_1^i\right)\ln\left(a\right)-\frac{1}{2\mu}\int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\left(\ln\left(a+\mu^i\right)-1\right)+1\right)d\mu^i}{a-1} \\ r_2^i &= \frac{A_2^i\ln\left(1+I\left(\mu^i\right)\right)-M\left(\mu^i\right)-I\left(\mu^i\right)}{I\left(\mu^i\right)}=\frac{\frac{1}{2}\left[\left(a+\mu^i+\varepsilon_2^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]}{a+\mu^i-1} \end{aligned}$$

We have already shown in the main text that $E\left(r_2^i|\mu^i\right)$ is positive and increasing in μ^i for $\mu^i>\mu^*$, and zero otherwise. To show that $E\left(r_2^i|r_1^i,\mu>\mu^i>\mu^*\right)$ is increasing in r_1^i, rewrite it as follows:

$$\begin{aligned} E\left(r_2^i|r_1^i,\mu>\mu^i>\mu^*\right) &= E_{\mu^i}\left(E\left(r_2^i|r_1^i,\mu^i,\mu>\mu^i>\mu^*\right)\right)=\int_{\mu^*}^{\mu}E\left(r_2^i|\mu^i\right)f\left(\mu^i|r_1^i,\mu>\mu^i>\mu^*\right)d\mu^i \\ &= \int_{\mu^*}^{\mu}\frac{\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]}{\left(a+\mu^i-1\right)}f\left(\mu^i|r_1^i,\mu>\mu^i>\mu^*\right)d\mu^i \end{aligned}$$

which implies

$$\frac{d}{dr_1^i}E\left(r_2^i|r_1^i,\mu>\mu^i>\mu^*\right)=\int_{\mu^*}^{\mu}\frac{\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]}{\left(a+\mu^i-1\right)}\frac{d}{dr_1^i}f\left(\mu^i|r_1^i,\mu>\mu^i>\mu^*\right)d\mu^i.$$

Since $\varepsilon_1^i|\mu^i\sim N\left(0,\sigma_{\varepsilon_1}^2\right)$ with $\sigma_{\varepsilon_1}^2=\frac{\sigma^2\left(I_0^i\right)^2}{\left[\ln\left(1+I_0^i\right)\right]^2}=\frac{\sigma^2(a-1)^2}{[\ln(a)]^2}$, we have $r_1^i|\mu^i,\mu>\mu^i>\mu^*\sim N\left(g\left(\mu^i\right),\sigma^2\right)$ with $g\left(\mu^i\right)=\frac{\mu^i\ln(a)-\frac{1}{2\mu}\int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\left(\ln\left(a+\mu^i\right)-1\right)+1\right)d\mu^i}{a-1}=\frac{\mu^i\ln(a)-x}{a-1}$, where $x=\frac{1}{2\mu}\int_{\mu^*}^{\mu}\left(\left(a+\mu^i\right)\left(\ln\left(a+\mu^i\right)-1\right)+1\right)d\mu^i$ does not depend on μ^i. Therefore,

$$\begin{aligned} f\left(r_1^i|\mu>\mu^i>\mu^*\right) &= \int_{\mu^*}^{\mu}f\left(r_1^i|\mu^i,\mu>\mu^i>\mu^*\right)f\left(\mu^i|\mu>\mu^i>\mu^*\right)d\mu^i \\ &= \int_{\mu^*}^{\mu}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}d\mu^i\frac{1}{\mu-\mu^*}=\frac{1}{\mu-\mu^*}\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right] \end{aligned}$$

and

$$\begin{aligned} f\left(\mu^i|r_1^i,\mu>\mu^i>\mu^*\right) &= f\left(r_1^i|\mu^i,\mu>\mu^i>\mu^*\right)\frac{f\left(\mu^i|\mu>\mu^i>\mu^*\right)}{f\left(r_1^i|\mu>\mu^i>\mu^*\right)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}\frac{\frac{1}{\mu-\mu^*}}{\frac{1}{\mu-\mu^*}\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right]}=\frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right]} \end{aligned}$$

for $\mu > \mu^i > \mu^*$, 0 otherwise, where ϕ and Φ are the pdf and cdf of the standard normal distribution, $z_{\mu^i} = \frac{r_1^i - g(\mu^i)}{\sigma}$, $z_\mu = \frac{r_1^i - g(\mu)}{\sigma}$, and $z_{\mu^*} = \frac{r_1^i - g(\mu^*)}{\sigma}$. Note that this simply says that $\mu^i | r_1^i, \mu > \mu^i > \mu^*$ is truncated normal, with truncation at $-\mu$ and μ^*. Since $\phi\left(z_{\mu^i}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}$, $\frac{d\phi\left(z_{\mu^i}\right)}{dr_1^i} = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} \frac{z_{\mu^i}}{\sigma} = -\phi\left(z_{\mu^i}\right) \frac{z_{\mu^i}}{\sigma}$ and

$$
\begin{aligned}
\frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i} &= \frac{-\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)\frac{z_{\mu^i}}{\sigma}}{\Phi\left(z_\mu\right) - \Phi\left(z_{\mu *}\right)} - \frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right) - \Phi\left(z_{\mu *}\right)\right]^2}\left[\phi\left(z_\mu\right)\left(\frac{1}{\sigma}\right) - \phi\left(z_{\mu^*}\right)\left(\frac{1}{\sigma}\right)\right] \\
&= \frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)\frac{1}{\sigma}}{\Phi\left(z_\mu\right) - \Phi\left(z_{\mu *}\right)}\left\{-z_{\mu^i} - \frac{\phi\left(z_\mu\right) - \phi\left(z_{\mu^*}\right)}{\Phi\left(z_\mu\right) - \Phi\left(z_{\mu *}\right)}\right\} \\
&= f\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)\frac{1}{\sigma}\left\{-z_{\mu^i} - \frac{\phi\left(z_\mu\right) - \phi\left(z_{\mu^*}\right)}{\Phi\left(z_\mu\right) - \Phi\left(z_{\mu *}\right)}\right\}.
\end{aligned}
$$

The expression $f\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)\frac{1}{\sigma}$ is positive for all values of μ^i. $\left\{-z_{\mu^i} - \frac{\phi(z_\mu) - \phi(z_{\mu^*})}{\Phi(z_\mu) - \Phi(z_{\mu *})}\right\}$ is increasing in μ^i (since $\frac{\phi(z_\mu) - \phi(z_{\mu^*})}{\Phi(z_\mu) - \Phi(z_{\mu *})}$ does not depend on i). Thus, there exists a value of μ^i, call it μ^x, which will depend on r_1^i and for which $\frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i} = 0$ for $\mu^i = \mu^x$, $\frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i} < 0$ for $\mu^i < \mu^x$, and $\frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i} > 0$ for $\mu^i > \mu^x$. Thus, $\frac{d}{dr_1^i} E\left(r_2^i | r_1^i, \mu > \mu^i > \mu^*\right) = \int_{\mu^*}^{\mu} E\left(r_2^i | \mu^i\right) \frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i} d\mu^i$ is positive (for all values of r_1^i) since $\int_{\mu^*}^{\mu} \frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i} d\mu^i = 0$ and $E\left(r_2^i | \mu^i\right)$ is positive and increasing, implying that the positive values of $\frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i}$ in $\int_{\mu^*}^{\mu} E\left(r_2^i | \mu^i\right) \frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i} d\mu^i$ are multiplied by a larger positive number than are the negative values of $\frac{df\left(\mu^i | r_1^i, \mu > \mu^i > \mu^*\right)}{dr_1^i}$.

Proof of Implication 1b

$$
\begin{aligned}
r_2^i &= \frac{A_2^i \ln\left(1 + I\left(\mu^i\right)\right) - M\left(\mu^i\right) - I\left(\mu^i\right)}{I\left(\mu^i\right)} = \frac{\frac{1}{2}\left[\left(a + \mu^i + \varepsilon_2^i\right)\ln\left(a + \mu^i\right) - \left(a + \mu^i - 1\right)\right]}{a + \mu^i - 1} \\
r_3^i &= \frac{A_3^i \ln\left(1 + I\left(\mu^i\right)\right) - M\left(\mu^i\right) - I\left(\mu^i\right)}{I\left(\mu^i\right)} = \frac{\frac{1}{2}\left[\left(a + \mu^i + \varepsilon_3^i\right)\ln\left(a + \mu^i\right) - \left(a + \mu^i - 1\right)\right]}{a + \mu^i - 1}
\end{aligned}
$$

We have already shown in the main text that $E\left(r_3^i | \mu^i\right)$ is positive and increasing in μ^i for $\mu^i > \mu^*$, and zero otherwise. To show that $E\left(r_3^i | r_2^i, \mu > \mu^i > \mu^*\right)$ is increasing in r_2^i, rewrite it as follows:

$$
\begin{aligned}
E\left(r_3^i | r_2^i, \mu > \mu^i > \mu^*\right) &= E_{\mu^i}\left(E\left(r_3^i | r_2^i, \mu^i, \mu > \mu^i > \mu^*\right)\right) = \int_{\mu^*}^{\mu} E\left(r_3^i | \mu^i\right) f\left(\mu^i | r_2^i, \mu > \mu^i > \mu^*\right) d\mu^i \\
&= \int_{\mu^*}^{\mu} \frac{\frac{1}{2}\left[\left(a + \mu^i\right)\ln\left(a + \mu^i\right) - \left(a + \mu^i - 1\right)\right]}{\left(a + \mu^i - 1\right)} f\left(\mu^i | r_2^i, \mu > \mu^i > \mu^*\right) d\mu^i
\end{aligned}
$$

which implies

$$
\frac{d}{dr_2^i} E\left(r_3^i | r_2^i, \mu > \mu^i > \mu^*\right) = \int_{\mu^*}^{\mu} \frac{\frac{1}{2}\left[\left(a + \mu^i\right)\ln\left(a + \mu^i\right) - \left(a + \mu^i - 1\right)\right]}{\left(a + \mu^i - 1\right)} \frac{d}{dr_2^i} f\left(\mu^i | r_2^i, \mu > \mu^i > \mu^*\right) d\mu^i.
$$

Since $\varepsilon_2^i|\mu^i \sim N\left(0, \sigma_{\varepsilon_2}^2\right)$ with $\sigma_{\varepsilon_2}^2 = \frac{\sigma^2\left(I_1^i\right)^2}{\left[\ln\left(1+I_1^i\right)\right]^2} = \frac{\sigma^2\left(a+\mu^I-1\right)^2}{[\ln(a+\mu^i)]^2}$, we have $r_2^i|\mu^i, \mu > \mu^i > \mu^* \sim N\left(h\left(\mu^i\right), \sigma^2\right)$ with $h\left(\mu^i\right) = \frac{\frac{1}{2}\left[\left(a+\mu^i\right)\ln\left(a+\mu^i\right)-\left(a+\mu^i-1\right)\right]}{a+\mu^i-1}$. We show in the main text that $h\left(\mu^i\right)$ is increasing in μ^i for $\mu^i > \mu^* = 1-a$.

Therefore,

$$\begin{aligned} f\left(r_2^i|\mu > \mu^i > \mu^*\right) &= \int_{\mu^*}^{\mu} f\left(r_2^i|\mu^i, \mu > \mu^i > \mu^*\right) f\left(\mu^i|\mu > \mu^i > \mu^*\right) d\mu^i \\ &= \int_{\mu^*}^{\mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} d\mu^i \frac{1}{\mu-\mu^*} = \frac{1}{\mu-\mu^*}\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right] \end{aligned}$$

and

$$\begin{aligned} f\left(\mu^i|r_2^i, \mu > \mu^i > \mu^*\right) &= f\left(r_2^i|\mu^i, \mu > \mu^i > \mu^*\right) \frac{f\left(\mu^i|\mu > \mu^i > \mu^*\right)}{f\left(r_2^i|\mu > \mu^i > \mu^*\right)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} \frac{\frac{1}{\mu-\mu^*}}{\frac{1}{\mu-\mu^*}\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right]} = \frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right]} \end{aligned}$$

for $\mu > \mu^i > \mu^*$, 0 otherwise, where ϕ and Φ are the pdf and cdf of the standard normal distribution, $z_{\mu^i} = \frac{r_2^i - h\left(\mu^i\right)}{\sigma}$, $z_\mu = \frac{r_2^i - h(\mu)}{\sigma}$, and $z_{\mu^*} = \frac{r_2^i - h(\mu^*)}{\sigma}$. Note that this just says that $\mu^i|r_2^i, \mu > \mu^i > \mu^*$ is truncated normal, with truncation at $-\mu$ and μ^*. Since $\phi\left(z_{\mu^i}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}$, $\frac{d\phi\left(z_{\mu^i}\right)}{dr_2^i} = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} \frac{z_{\mu^i}}{\sigma} = -\phi\left(z_{\mu^i}\right)\frac{z_{\mu^i}}{\sigma}$ and

$$\begin{aligned} \frac{df\left(\mu^i|r_2^i, \mu > \mu^i > \mu^*\right)}{dr_2^i} &= \frac{-\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)\frac{z_{\mu^i}}{\sigma}}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)} - \frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)\right]^2}\left[\phi\left(z_\mu\right)\left(\frac{1}{\sigma}\right)-\phi\left(z_{\mu^*}\right)\left(\frac{1}{\sigma}\right)\right] \\ &= \frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)\frac{1}{\sigma}}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)}\left\{-z_{\mu^i} - \frac{\phi\left(z_\mu\right)-\phi\left(z_{\mu^*}\right)}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)}\right\} \\ &= f\left(\mu^i|r_2^i, \mu > \mu^i > \mu^*\right)\frac{1}{\sigma}\left\{-z_{\mu^i} - \frac{\phi\left(z_\mu\right)-\phi\left(z_{\mu^*}\right)}{\Phi\left(z_\mu\right)-\Phi\left(z_{\mu*}\right)}\right\}. \end{aligned}$$

The expression $f\left(\mu^i|r_2^i, \mu > \mu^i > \mu^*\right)\frac{1}{\sigma}$ is positive for all values of μ^i. $\left\{-z_{\mu^i} - \frac{\phi(z_\mu)-\phi\left(z_{\mu^*}\right)}{\Phi(z_\mu)-\Phi(z_{\mu*})}\right\}$ is increasing in μ^i (since $\frac{\phi(z_\mu)-\phi\left(z_{\mu^*}\right)}{\Phi(z_\mu)-\Phi(z_{\mu*})}$ does not depend on i). Thus, there exists a value of μ^i, call it μ^{xx} (which will depend on r_2^i) for which $\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_2^i} = 0$ for $\mu^i = \mu^{xx}$, $\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_1^i} < 0$ for $\mu^i < \mu^{xx}$, and $\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_2^i} > 0$ for $\mu^i > \mu^{xx}$. Therefore, $\frac{d}{dr_2^i}E\left(r_3^i|r_2^i, \mu > \mu^i > \mu^*\right) = \int_{\mu^*}^{\mu} E\left(r_3^i|\mu^i\right)\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_2^i}d\mu^i$ is positive (for all values of r_2^i) since $\int_{\mu^*}^{\mu}\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_2^i}d\mu^i = 0$ and $E\left(r_3^i|\mu^i\right)$ is positive and increasing, implying that in $\int_{\mu^*}^{\mu} E\left(r_3^i|\mu^i\right)\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_2^i}d\mu^i$ the positive values of $\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_2^i}$ are multiplied by a larger positive number than are the negative values of $\frac{df\left(\mu^i|r_2^i,\mu>\mu^i>\mu^*\right)}{dr_2^i}$.

Proof of Implication 1c

We have verified numerically that persistence is higher from second to third funds than from first to second funds, for all values of σ^2, μ and a (with $a > 1$). We do this by constructing a data set consistent with the model as follows: We define a vector of equally spaced values of μ^i between μ^* and μ. For each value of μ^i we draw values for ε_1^i, ε_2^i and ε_3^i and define r_1^i, r_2^i, and r_3^i. We then regress r_2 on r_1 (and a constant) and regressed r_3 on r_2 (and a constant) and compare the regression coefficients. We do this for for a wide range of parameter values σ^2, μ and $a > 1$.