RFS Advance Access published July 31, 2013

Informational Holdup and Performance Persistence in Venture Capital

Yael V. Hochberg
Northwestern University and NBER

Alexander Ljungqvist
New York University, CEPR, and NBER

Annette Vissing-Jørgensen
University of California, Berkeley, CEPR, and NBER

Why don't VCs eliminate excess demand for follow-on funds by raising fees? We propose a model of learning that leads to informational holdup. Current investors learn about skill whereas outside investors observe only returns. This gives current investors holdup power when the VC raises his next fund: Without their backing, no-one will fund him, as outside investors interpret the lack of backing as a sign of low skill. Holdup power diminishes the VC's ability to increase fees in line with performance, leading to return persistence. Empirical evidence supports the model. We estimate that up to two-thirds of VC firms lack skill. (*JEL* G24)

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 allocations in their next fund to the highest bidder?

As Berk and Green (2004) show in the context of mutual funds, if capital supply is competitive but fund management skill is scarce, investors earn zero expected excess returns net of fees, realized returns are unpredictable from public information, and fund managers earn fees reflecting their skill. This fits

We are grateful to Viral Acharya, Ulf Axelson, April Franco, Paolo Fulghieri (the Editor), Rick Green, Arvind Krishnamurthy, Adair Morse, Ludovic Phalippou, Morten Sørensen, Chris Yung, an anonymous reviewer, and to seminar participants at the WFA meetings, the NBER, Queen's University, 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. Send correspondence to Alexander Ljungqvist, Stern School of Business, New York University, 44 W. 4th St. #9-160, NY, NY 10012, USA; telephone: 212-998-0304. E-mail: aljungqv@stern.nyu.edu.

doi:10.1093/rfs/hht046

mutual funds, whose 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. Instead, 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 holdup designed to explain 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 interim 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 that they hold for a number of years before eventually selling (or, more often, writing off). Objective returns thus take many years to materialize, unlike in the mutual fund industry, where managers invest in traded securities that can be easily and objectively 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, his skill is unknown and so he faces a large set of potential investors in a perfectly

[1] For empirical evidence of the importance of soft information in learning about corporate managers' skill, see Cornelli, Kominek, and Ljungqvist (2013).

[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 to inform their reinvestment decision. Chung et al. (2012) use a learning model to calibrate the incentive effects of future fundraising in the VC market.

competitive LP market. But over time, as "incumbent" LPs discover his skill before outside investors do, the LP market becomes less competitive. This asymmetric learning in turn enables incumbent LPs to hold up the GP 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 holdup problem prevents the GP from raising the fee to the point at which 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 potentially 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 up the GP, 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 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 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 dispersion 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).

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

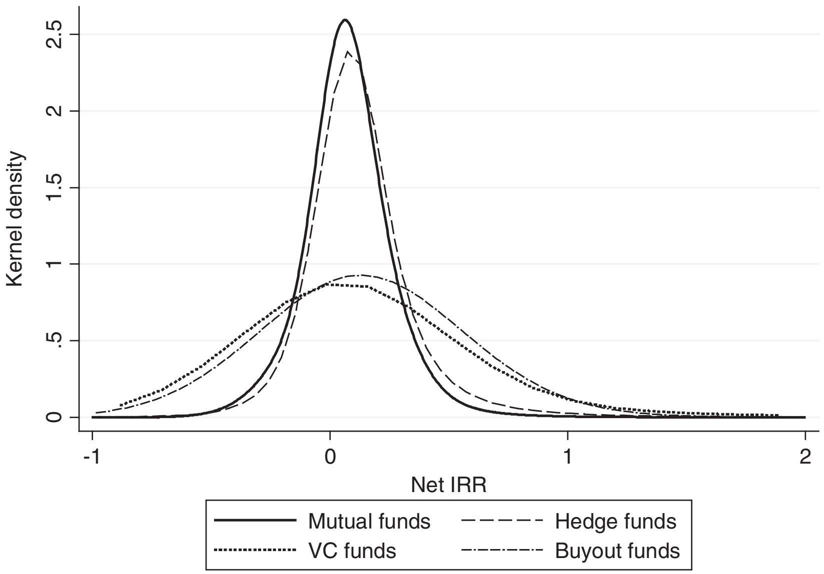

Figure 1
Fund risk
This figure shows the distribution of net IRRs for mutual funds, hedge funds, VC funds, and buyout funds in the United States for the period from 1980 to 2006. The graph presents, for each set of funds, Gaussian kernel densities of net annual IRRs from CRSP (for mutual funds), the CISDM Hedge Funds database available on WRDS (for hedge funds), PREQIN (for buyout funds), and a 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 ten 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.

distinction between soft and hard information that allows us to test whether informational holdup can explain performance persistence in venture capital.

We test the model's predictions using both survey and observational data. The former come from a survey of LPs in buyout and VC funds conducted by Marco Da Rin and Ludovic Phalippou.[3] The questionnaire includes the question, "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, 87.5% indicate receiving priority over outside investors in follow-on funds. Moreover, 72.1% of these LPs agree with the following statement: "If I did not 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

[3] For further details of the survey and its methodology, see Da Rin and Phalippou (2011).

Table 1
Survey evidence: Do LPs receive priority in follow-on funds, and if so, why?

	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 did not reinvest, 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 did not 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

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 similar if only the responses of U.S.-based LPs are tabulated.

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, that is, before their current fund's final return is known. At this point, the market only knows 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.[4] Incumbent LPs observe the reported interim return, 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

4 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.

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 asymmetric learning and so with the economic mechanism at the heart of our model—informational holdup.

Finally, our data allow us to estimate the prevalence of skill in the VC industry. The model predicts that VCs go out of business (in the sense of being unable to raise a follow-on fund) once their LPs 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 literatures on venture capital, holdup in banking relationships, and learning in financial markets. In contrast to the banking setting, asymmetric learning is efficient in venture capital, in the sense that VCs raise the first-best amounts of capital in both first and follow-on funds. Moreover, GPs may even prefer informational holdup 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 be communicated credibly to potential new investors.

We proceed as follows. Section 1 presents our model of learning and informational holdup. Sections 2 and 3 present the data and empirical analysis. Section 4 concludes.

1. A Model of Learning about GP Skill

To provide intuition for our core result that learning and informational holdup generate persistence and to show that this result does not rely on a particular bargaining mechanism, we begin with a simple example. The example simplifies our model by assuming that GPs and LPs are risk neutral, each fund has a single LP and lasts one period, fund sizes (but not fees) are exogenous, and all GPs raise both a first fund and a follow-on fund. Our main model relaxes each of these assumptions.

1.1 Example

Consider a setting in which each fund has a single GP. At $t=0$ and again at $t=1$, a GP raises a fund of size I, lasting one period each. GPs differ in skill. Skill determines the properties of a fund's cash flows and is captured by μ^i. For GP i and fund $k \in \{1,2\}$, the cash flows are $C_k^i = A_k^i I$, where $A_k^i = a + \mu^i + \varepsilon_k^i$ with $\varepsilon_k^i | \mu^i \sim N(0, \sigma^2)$. ε_1^i, ε_2^i are independent. 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.

There is a large set of ex ante identical investors. Each fund has one LP. Both GPs and LPs are risk neutral and consume at $t=2$. GPs and LPs have initial wealth of W_0^{GP} and W_0^{LP}, respectively. Besides investing in venture capital, LPs can invest at a risk-free rate of r_f, set equal to zero without loss of generality. We assume that LPs can invest in one first fund and, if desired, in the GP's second fund. Cash flows received at $t=1$ are invested at rate r_f until $t=2$.

At time $t=0$, no one knows the GP's skill. At $t=1$, the GP and the LP who invested in the GP's first fund learn the GP's skill. Investors who have not invested in a GP's first fund only observe its cash flow (and fund size, I). 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.

We assume that a fund's cash flow is divided between the GP and LP according to the following contract, agreed at the start of the fund. In a first fund, GP i receives a dollar amount of $X_1^{GP} = M_1$ at $t=1$, while the LP receives $X_1^{LP} = C_1^i - M_1$. In a second fund, payoffs are $X_2^{GP} = M_2$ and $X_2^{LP} = C_2^i - M_2$, received at $t=2$. LP returns (after fees) are thus

$$r_1^i = A_1^i - \frac{M_1}{I} - 1 = a + \mu^i + \varepsilon_1^i - \frac{M_1}{I} - 1 \text{ for the first fund,}$$

$$r_2^i = A_2^i - \frac{M_2(\mu^i)}{I} - 1 = a + \mu^i + \varepsilon_2^i - \frac{M_2(\mu^i)}{I} - 1 \text{ for the second fund.}$$

We are interested in how the division of the second fund's cash flow determines whether we observe persistence. Starting with first funds, as no learning has taken place yet, the LP market is perfectly competitive at $t=0$. Accordingly, LPs have no bargaining power and all GPs offer a contract that ensures the highest possible expected payoff for the GP, subject to each LP achieving an expected payoff (across investing in both the GP's first and second funds) equal to the outside option (which is I for each fund given $r_f=0$). We refer to this as the LP's participation constraint,

$$E_{\mu^i}\left[\left(E\left(A_1^i|\mu^i\right)I - M_1 - I\right) + \left(E\left(A_2^i|\mu^i\right)I - M_2\left(\mu^i\right) - I\right)\right] = 0$$

$$\Longleftrightarrow M_1 = 2(a-1)I - E_{\mu^i}\left(M_2\left(\mu^i\right)\right).$$

With 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 skill, the incumbent LP has an informational advantage when the GP raises a second fund and so can extract part of the surplus. For now, we simply assume that outside investors cannot participate in second funds. The complete model lays out a formal bargaining process to derive a winner's curse, which causes outside investors to rationally abstain from participating in second funds when given the chance. With no outside investors to rely on, the GP and the incumbent LP share the second-fund surplus. The following result, which we prove in the Appendix, states a sufficient condition for performance persistence.

Result 1: Performance persistence. If $E\left(r_2^i|\mu^i\right)$ increases in GP skill μ^i, which is the case for $\frac{M_2'(\mu^i)}{I}<1$, then a high after-fee return to the LP in a GP's first fund predicts a high after-fee return in the second fund: $E\left(r_2^i|r_1^i\right)$ increases in r_1^i.

The condition $\frac{M_2'(\mu^i)}{I}<1$ states that the GP's fee as a percent of invested capital, $\frac{M_2(\mu^i)}{I}$, needs to increase less quickly in GP skill than the fund's expected pre-fee return (which equals $E\left(A_2^i\right)-1=a+\mu^i-1$ and so increases one for one with μ^i). This is satisfied in all the most commonly used approaches to bargaining and so our core results do not rely on any particular form of bargaining over the fund surplus. We show this for three popular forms of bargaining: Nash bargaining, Shapley values and Binmore-Rubinstein-Wolinsky alternating offers.

Suppose the GP and the incumbent LP have bargaining power $g\left(\mu^i\right)$ and $1-g\left(\mu^i\right)$, respectively, with $0<g\left(\mu^i\right)<1$. Then the Nash bargaining outcome is that the expected surplus $\left[E\left(A_2|\mu^i\right)-1\right]I$ is shared such that the GP receives an expected payoff of $M_2\left(\mu^i\right)=g\left(\mu^i\right)\left[E\left(A_2|\mu^i\right)-1\right]I=g\left(\mu^i\right)\left[a+\mu^i-1\right]I$ and the LP receives an expected payoff of $\left(1-g\left(\mu^i\right)\right)\left[E\left(A_2|\mu^i\right)-1\right]I$. This implies that $M_2'\left(\mu^i\right)=\left(g\left(\mu^i\right)+g'\left(\mu^i\right)\left[a+\mu^i-1\right]\right)I$. The condition $\frac{M_2'(\mu^i)}{I}<1$ is satisfied if bargaining power is constant and so unrelated to GP skill (i.e., $g'\left(\mu^i\right)=0$ and $M_2'\left(\mu^i\right)=g\left(\mu^i\right)I<I$). It is also satisfied if GP bargaining power is increasing in skill, as long as $g'\left(\mu^i\right)$ is not too large.

Shapley values lead to the same outcome as that resulting from Nash bargaining with $g=0.5$. The expected cash flow (net of the investment) of the follow-on fund is $\left[E\left(A_2|\mu^i\right)-1\right]I$, whereas neither the GP nor LP obtains any payoff if the other party does not participate in the follow-on fund. Thus, the Shapley value to each party is $\frac{1}{2}\left[E\left(A_2|\mu^i\right)-1\right]I$.

Finally, Binmore-Rubinstein-Wolinsky alternating offers also lead to the same outcome as Nash bargaining with $g=0.5$. The GP and LP take turns making offers with a particular value for $M_2\left(\mu^i\right)$. In between each offer, there is an exogenous risk that the bargaining process will terminate without agreement. Binmore, Rubinstein, and Wolinsky (1986) show that the equilibrium outcome is immediate agreement with the first offer being accepted and that the payoffs converge to those for symmetric Nash bargaining as the probability of exogenous breakdown goes to zero.

1.2 Main model

We now turn to our main model. Its primary objective is to show why the presence of multiple LPs does not result in "Bertrand competition" such that the GP has all the bargaining power.

1.2.1 Setup. *Fund cash flow:* 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 $[-\mu,\mu]$; $\mu^i=0$ corresponds to average skill. At this time, no one knows which GPs have skill. We abstract from agency problems by assuming that GPs act in their LPs' best interest.[5] A first fund has size I_0, lasts two periods, and generates cash flows that are paid out to LPs at $t=2$, the end of the second period.

Between $t=0$ and $t=1$, the GP invests in portfolio companies whose performance is hit by a random shock, ε^i. 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 $C_2^i = A_2^i \ln\left(1+I_0^i\right)$.[6] A_2^i captures the effects of the two random shocks and the GP's skill, μ^i:

$$A_2^i = a + H_1(\mu^i,\varepsilon^i) + H_2(\mu^i,v^i), \tag{1}$$

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

We assume $H_1(\mu^i,\varepsilon^i)=\mu^i+\varepsilon^i$ and $H_2(\mu^i,v^i)=\mu^i+v^i$.

Depending on what is learned between $t=0$ and $t=1$ (described below), 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, that is, before its final IRR is publicly known.[7] 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

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

[6] The log function captures decreasing returns to scale. This is similar to Berk and Green's (2004) assumption for mutual funds.

[7] Our two-fund setup is intended to capture the two key "periods" in institutional reality—the period in which everyone lacks good information about the VC's true skill (the "first fund") and a subsequent period when insiders, due to their access as LPs, have managed to learn the VC's true skill through gathering soft information (the "follow-on fund"). In practice, a VC may need to raise more than one "first fund" before his skill is revealed.

$t=3$. Using subscript "follow-on" for such funds, we assume that

$$A_3^i = a + H_{2,follow-on}\left(\mu^i, \varepsilon^i_{follow-on}\right) + H_{3,follow-on}\left(\mu^i, v^i_{follow-on}\right), \quad (3)$$

$$H^i_{2,follow-on} = \mu^i + \varepsilon^i_{follow-on}, \; H^i_{3,follow-on} = \mu^i + v^i_{follow-on}, \quad (4)$$

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

where $\varepsilon^i_{follow-on}$ and $v^i_{follow-on}$ are independent of each other and of ε^i and v^i. All shocks are independent across GPs and thus of skill, μ^i.[8] 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 LPs will not eliminate the informational holdup that is at the heart of our model, while still preserving mathematical tractability.[9] GPs do not invest in their own funds.[10] At $t=1$, we distinguish between incumbent LPs, who have invested in a given GP's first 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 it is noisily reflected in the fund's interim performance. Let $H_1^i = H_1(\mu^i, \varepsilon^i) = \mu^i + \varepsilon^i$ be the hard information about interim performance that outside investors can verify at $t=1$. The challenge for outside investors is to disentangle skill μ^i from luck ε^i. Based on observing H_1^i, outside investors update their beliefs about the GP's skill from the unconditional mean of $E\left(\mu^i\right)=0$ to $E\left(\mu^i | H_1^i\right)$. The hard information available at $t=1$ can be thought of as (monotonically increasing in) the fund's interim return 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. At $t=2$, the first fund's final cash flow (which is monotonically increasing in its final return) becomes public information.

---

8 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.

9 While modeling the optimal number of LPs would make the model intractable, the intuition for why multiple informed LPs do not compete away their holdup power does not, as we will show, depend on the number of LPs.

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

We refer to this setup as asymmetric learning, in the sense that incumbent LPs learn the GP's type faster than do outside investors.[11] 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 risk-free 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 funds are invested at the risk-free 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 $X_0^{GP}=M_{0,a}+M_{0,b}\equiv 2M_0$, whereas the two LPs each receive cash flows net of fees equal to $X_0^{LP}=C_2^i/2-M_0$.[12]

As we will see shortly, follow-on funds have either one or two LPs. If both incumbents 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 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 are 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.

1.2.2 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 allows them to extract part of the follow-on fund's value from the GP.

Although 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 holdup that is our

[11] 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.

[12] 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.

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 Binmore, Rubinstein, and Wolinsky (1986) 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:

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

(2) 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 $M_{1,split}^{LP_a}$ or to provide all of the capital needed (a sole offer) and pay a fee of $M_{1,sole}^{LP_a}$. This offer is denoted $\left[(I_{1,split}^{LP_a}/2, M_{1,split}^{LP_a}), (I_{1,sole}^{LP_a}, M_{1,sole}^{LP_a})\right]$. Similarly, LP_b's offer is $\left[(I_{1,split}^{LP_b}/2, M_{1,split}^{LP_b}), (I_{1,sole}^{LP_b}, M_{1,sole}^{LP_b})\right]$.[13] The GP can accept either both split offers or one of the sole offers or reject both offers.

(3) 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 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 terminates without agreement.[14]

If no agreement is reached, each party receives its outside option. For the incumbent LPs, this equals the risk-free 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 one of the parties has simply refused to bargain further. We also 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 their expected utility. Why? If the outside offer resulted in a gain in expected utility for the investor who made

[13] 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, 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.

[14] 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.

it, it would be immediately 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 skill was sufficiently low so that the incumbents chose not to counter. Outside 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 fund sizes that maximize the joint surplus of the GP and LPs in the split and sole cases. We use superscript i to denote their dependence on the GP's skill, μ^i: $I^i_{1,split}$ and $I^i_{1,sole}$. Furthermore, at $t=1$ the GP and incumbent LPs know μ^i and ε^i.

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

$$\max_{I_1} E\left(U^{GP}|\mu^i,\varepsilon^i\right)+2E\left(U^{LP}|\mu^i,\varepsilon^i\right), \tag{5}$$

where

$$E\left(U^{GP}|\mu^i,\varepsilon^i\right)=1-e^{-\gamma W_3^{GP}}=1-e^{-\gamma\left[W_0^{GP}+2M_0+2M_1\right]} \tag{6}$$

$$\begin{aligned} E\left(U^{LP}|\mu^i,\varepsilon^i\right)&=1-E\left(e^{-\gamma W_3^{LP}}|\mu^i,\varepsilon^i\right)\\ &=1-e^{-\gamma\left[W_0^{LP}-M_0-M_1\right]}E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]}|\mu^i,\varepsilon^i\right)\\ &\quad E\left(e^{-\gamma\frac{1}{2}[A_3\ln(1+I_1)-I_1]}|\mu^i\right). \end{aligned} \tag{7}$$

Only $E\left(e^{-\gamma\frac{1}{2}[A_3\ln(1+I_1)-I_1]}|\mu^i\right)$ depends on I_1 and because A_3 is normally distributed with $V\left(A_3|\mu^i\right)=\frac{\sigma^2(I_1)^2}{[\ln(1+I_1)]^2}$, we obtain $E\left(e^{-\gamma\frac{1}{2}[A_3\ln(1+I_1)-I_1]}|\mu^i\right)=e^{-\gamma\left[\frac{1}{2}\left(E\left(A_3|\mu^i\right)\ln(1+I_1)-I_1\right)-\frac{1}{8}\gamma\sigma^2(I_1)^2\right]}$, which is maximized by

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

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

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

$$\max_{I_1} E\left(U^{GP}|\mu^i,\varepsilon^i\right)+E\left(U^{LP}|\mu^i,\varepsilon^i\right). \tag{10}$$

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

$$I^i_{1,sole} = \frac{E\left(A_3|\mu^i\right)}{1+\gamma\sigma^2 I^i_{1,sole}}-1 \quad (11)$$

$$= \frac{-\left(1+\gamma\sigma^2\right)+\sqrt{\left(1+\gamma\sigma^2\right)^2+4\gamma\sigma^2\left[E\left(A_3|\mu^i\right)-1\right]}}{2\gamma\sigma^2}. \quad (12)$$

Both $I^i_{1,split}$ and $I^i_{1,sole}$ equal zero for $E\left(A_3|\mu^i\right)=1$. Because $E\left(A_3|\mu^i\right)=a+2\mu^i$, 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^*.

Discussion: Whatever the number of LPs in the follow-on fund, the optimal fund size does not depend on the fee, M_1. Instead, it maximizes the LPs' risk-adjusted cash flows, and 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 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, whatever 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.

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 an LP's risk-adjusted (pre-fee, net of investment) cash flows in the case of a split and sole outcome, respectively, as

$$b_{split}\left(\mu^i\right) = \frac{1}{2}\left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right]-\frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2 \quad (13)$$

$$b_{sole}\left(\mu^i\right) = \left[E\left(A_3|\mu^i\right)\ln\left(1+I^i_{1,sole}\right)-I^i_{1,sole}\right]-\frac{1}{2}\gamma\sigma^2\left(I^i_{1,sole}\right)^2, \quad (14)$$

and define the fee function $M^*_1\left(\mu^i\right)=g\left(\mu^i\right)b_{split}\left(\mu^i\right)$ with $g\left(\mu^i\right)=\frac{-\ln\left(x\left(\mu^i\right)\right)}{\gamma b_{split}\left(\mu^i\right)}$, where $x\left(\mu^i\right)$ is the real root of the cubic equation $2e^{\gamma b_{split}\left(\mu^i\right)}x(\mu_i)^3-x(\mu_i)^2=0$ (as derived in the Appendix). Then $g\left(\mu^i\right)$ is monotonically decreasing in μ^i, $g\left(\mu^i\right)\to 1/2$ as $b_{split}\left(\mu^i\right)\to 0$, and $g\left(\mu^i\right)\to 1/3$ as $b_{split}\left(\mu^i\right)\to\infty$. Furthermore, provided that

$$b_{split}\left(\mu^i\right)-M^*_1\left(\mu^i\right) > b_{sole}\left(\mu^i\right)-2M^*_1\left(\mu^i\right) \quad (15)$$

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 strictly 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 (1) 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 (2) 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 the Appendix.[15]

Corollary 1. For given skill μ^i and thus $E\left(A_3|\mu^i\right)$, condition (15) in Proposition 1 is satisfied for γ and σ^2 sufficiently high. Specifically, for 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 condition (15) is satisfied for $\gamma > \gamma^*\left(\sigma^2, E\left(A_3|\mu^i\right)\right)$. Moreover, $\gamma^*\left(\sigma^2, E\left(A_3|\mu^i\right)\right)$ 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 $E\left(A_3|\mu^i\right)$.

Corollary 2. For $\mu^i > \mu^*$, the LPs' risk-adjusted cash flows in follow-on funds after fees, $\left(1-g\left(\mu^i\right)\right) b_{split}\left(\mu^i\right)$, are positive and increasing in the GP's skill, μ^i.

15 While the Appendix 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.

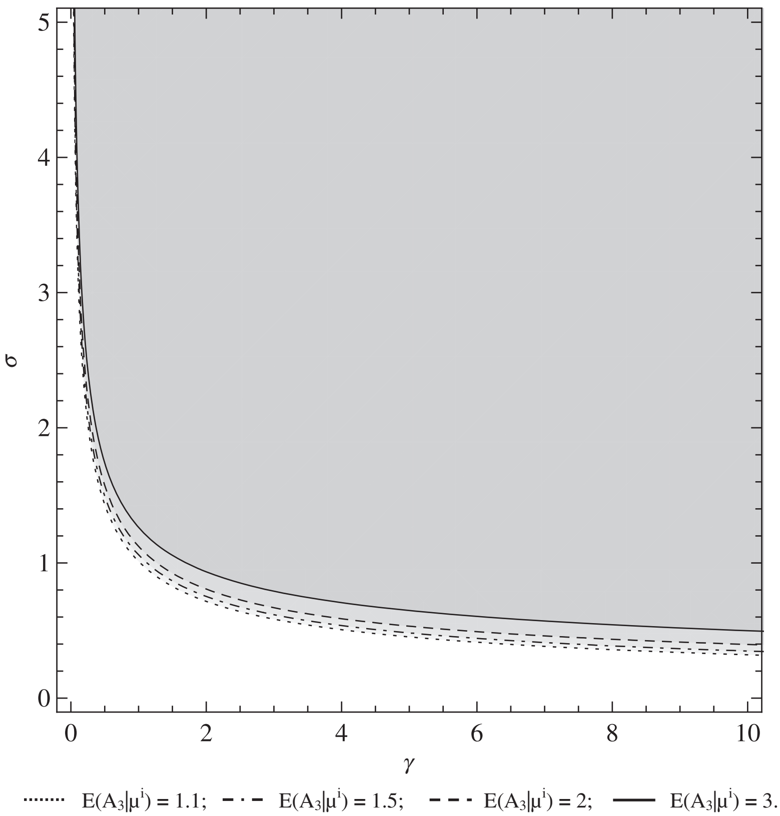

Figure 2
Illustration of Corollary 1
This figure illustrates Corollary 1 by depicting the values of γ and σ (shaded area) for which the condition in Proposition 1 holds, for values of $E(A_3|\mu^i)$ in the set [1.1 1.5 2 3].

Proof. See the Appendix. ■

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.

Discussion: Our bargaining game is conceptually harder to solve than standard Rubinstein (1982) bargaining, as there are three risk-averse parties involved (one GP, two LPs). With two LPs, one has to consider if the LPs will play the (appropriately adapted version of the) standard Rubinstein (1982) strategy or try to outbid each other to obtain a larger fund allocation. Corollary 1 shows that for sufficiently high risk aversion and risk, LPs will play the standard Rubinstein (1982) strategy. The intuition is straightforward. By slightly undercutting the other LP, an LP could become the sole investor in the fund. This benefits the

LP because in the Rubinstein (1982) solution, part of the fund surplus goes to the LPs, and receiving the full LP surplus is better than getting only part of it. But this outcome has two types of costs: First, idiosyncratic risk becomes increasingly costly (in utility terms) the more an LP invests in a fund. Second, the LP pays the fee on the entire fund, rather than splitting the cost with the second LP. Condition (15) is hence 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 (15) reflects this.

Does Proposition 1 generalize to other bargaining games? Shapley values cannot be applied in settings with risk averse players. The Appendix shows that symmetric Nash bargaining results in the GP receiving a fee equal to one-third of the LPs' risk-adjusted cash flow, $M_1\left(\mu^i\right)=\frac{1}{3}b_{split}\left(\mu^i\right)$. This in turn implies performance persistence, as Corollary 2 still holds with $g\left(\mu^i\right)$ set to 1/3.

1.2.3 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 (15) in Proposition 1 holds.

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

$$W_3^{LP}=W_0^{LP}+\frac{1}{2}(A_2\ln(1+I_0)-I_0)$$
$$+\frac{1}{2}\left[A_3\ln\left(1+I_{1,split}^i\right)-I_{1,split}^i\right]-M_0-M_1^*\left(\mu^i\right) \quad (16)$$

and

$$W_3^{GP}=W_0^{GP}+2M_0+2M_1^*\left(\mu^i\right). \quad (17)$$

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: At $t=0$, 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. Define

$$c\left(\mu^i\right)=\frac{1}{2}\left[E\left(A_2|\mu^i\right)\ln(1+I_0)-I_0\right]-\frac{1}{8}\gamma\sigma^2(I_0)^2 \quad (18)$$

Then

$$EU^{LP}=E_{\mu^i}\left(E\left(1-e^{-\gamma W_3^{LP}}|\mu^i\right)\right)$$
$$=1-e^{-\gamma W_0^{LP}}E_{\mu^i}\left(e^{-\gamma\left[c(\mu^i)-M_0\right]}e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}\right), \quad (19)$$

exploiting that A_2 and A_3 are uncorrelated conditional on μ^i. The LPs' participation constraint is that $EU^{LP}=1-e^{-\gamma W_0^{LP}}$, that is, that

$$M_0(I_0)=-\frac{1}{\gamma}\ln E_{\mu^i}\left(e^{-\gamma c(\mu^i)}e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}\right), \quad (20)$$

where $b_{split}(\mu^i)=0$ for $\mu^i<\mu^*$. The participation constraint simply says that, to the extent that LPs (due to their informational holdup power) earn a positive risk-adjusted cash flow after fees in follow-on funds ($b_{split}(\mu^i)-M_1^*(\mu^i)>0$), first 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:

$$\max_{I_0} E_{\mu^i}\left(1-e^{-\gamma W_3^{GP}}\right) \text{ s.t. } M_0=M_0(I_0). \quad (21)$$

As $E_{\mu^i}\left(1-e^{-\gamma W_3^{GP}}\right)=1-e^{-\gamma W_0^{GP}}e^{-\gamma 2M_0}E_{\mu^i}\left(e^{-\gamma 2M_1^*(\mu^i)}\right)$, and because $M_1^*(\mu^i)$ from Proposition 1 does not depend on what happens in the first fund, this implies simply choosing the value of I_0 that maximizes $M_0(I_0)$.[16] Because

$$E_{\mu^i}\left(e^{-\gamma c(\mu^i)}e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}\right)=\frac{1}{2\mu}\int_{-\mu}^{\mu}e^{-\gamma c(\mu^i)}d\mu^i$$
$$+\frac{1}{2\mu}\int_{\mu^*}^{\mu}e^{-\gamma c(\mu^i)}\left[e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}-1\right]d\mu^i, \quad (23)$$

[16] This is equivalent to choosing I_0 to maximize the joint surplus without constraints, because

$$EU^{GP}+2EU^{LP}$$
$$=1-e^{-\gamma W_0^{GP}}e^{-\gamma 2M_0}E_{\mu^i}\left(e^{-\gamma 2M_1^*(\mu^i)}\right)$$
$$+2\left[1-e^{-\gamma W_0^{LP}}e^{\gamma M_0}E_{\mu^i}\left(e^{-\gamma c(\mu^i)}e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}\right)\right], \quad (22)$$

of which only $E_{\mu^i}\left(e^{-\gamma c(\mu^i)}e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}\right)$ depends on I_0 (via $c(\mu^i)$).

the first-order condition for the optimal first-fund size, I_0, that maximizes $M_0(I_0)$ is

$$0=\int_{-\mu}^{\mu} e^{-\gamma c(\mu^i)}\left\{\left(\frac{E\left(A_2|\mu^i\right)}{(1+I_0)}-1\right)-\gamma\frac{1}{2}\sigma^2 I_0\right\}d\mu^i+$$
$$\int_{\mu^*}^{\mu} e^{-\gamma c(\mu^i)}\left\{\left(\frac{E\left(A_2|\mu^i\right)}{(1+I_0)}-1\right)-\gamma\frac{1}{2}\sigma^2 I_0\right\}\left[e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}-1\right]d\mu^i. \tag{24}$$

The first integral captures the optimal first-fund size, considering it 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. Given Proposition 1, $b_{split}\left(\mu^i\right)-M_1^*\left(\mu^i\right)>0$ for $\mu^i>\mu^*$, so $\left[e^{-\gamma\left[b_{split}(\mu^i)-M_1^*(\mu^i)\right]}-1\right]<0$ for $\mu^i>\mu^*$. So 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. Intuitively, the marginal value of increasing fund size I_0 is reduced by the fact that risk-adjusted cash flows are unconditionally (absent knowledge of μ^i at $t=0$) positively correlated across a GP's two funds.

First-fund fee: Whereas the optimal size of a first fund, I_0^*, cannot be derived in closed form, its fee, for any given I_0, can be determined directly from $M_0(I_0)$.

1.3 Empirical predictions

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

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

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

and let $r^i_{follow-on,final}$ denote the realized after-fee return in GP i's follow-on fund at $t=3$,

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

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

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

where we omit a superscript i on I_0 as 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: $E(r^i_{follow-on,final} | r^i_{first,interim}, \mu^i > \mu^*) - E(\frac{1}{4}\gamma\sigma^2 I^i_{1,split} | r^i_{first,interim}, \mu^i > \mu^*)$ increases in $r^i_{first,interim}$.[17]

Implication 1 is, of course, what the model is designed to capture. The proof is presented in the Appendix. 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.

Our model assumes that each GP raises at most two funds. In practice, GPs often raise more than two funds over time. Would our model predict persistence in later funds? The answer is yes, as long as incumbent LPs continue to have an informational advantage over outside investors and so retain some bargaining power over the GP.[18] This will be the case as long as the incumbent LPs have not fully learned the GP's skill; once they have, fund size will be stable over time and reveal GP skill to outside investors who then can compete with incumbent LPs in subsequent funds. This competition would result in returns equal to the LPs' outside option, and persistence would come to an end.

It is an empirical question when the incumbent LPs learn the GP's skill. Our two-fund model assumes that they have done so by the time the follow-on fund is raised, such that the size of the follow-on fund, once raised, reveals the GP's skill from then on. In practice, as we argue more fully in Section 2.3, learning

17 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 $E\left(r^i_{follow-on,final} | r^i_{first,interim}, \mu^i > \mu^*\right) - E\left(\gamma \frac{1}{4}\sigma^2 I^i_{1,split} | r^i_{first,interim}, \mu^i > \mu^*\right)$.

18 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.

may take considerably longer. In the limit, if incumbent LPs, though learning from fund to fund, never fully discover the GP's skill, persistence will persist forever.[19]

1.3.2 Additional empirical implications. In addition to performance persistence, the model yields empirical implications concerning the probability that a follow-on fund is raised, its size, and its expected return.

Implications 2a and 3a concern the impact of learning on fundraising and fund size and hold whether learning is symmetric or asymmetric. (We solve the model with symmetric learning in the proof of Implication 2a in the Appendix.) 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, whereas $r^i_{first,final}$ serves as a proxy for the incumbent LPs' soft information about GP skill (i.e., their knowledge of μ^i).

Implication 2: Fundraising
(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:

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

(2) 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 predicts follow-on fundraising, 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:

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

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

[19] An alternative version of informational dynamics which retains persistence beyond second funds involves the (realistic) possibility that a GP's skill may change gradually over time. This could happen if, for example, the GP team changes over time as some general partners retire and new ones are hired. Then as long as incumbents are always better informed about the GP team's current skill, outside investors will choose not to participate in third and later funds so that the incumbent LPs retain their informational holdup over the GP.

(b) If learning is asymmetric, soft information predicts follow-on fund size, over and above the hard information available to outside investors: $E\left(I_1^i | r^i_{first,interim}, r^i_{first,final}, \mu^i > \mu^*\right)$ increases in $r^i_{first,final}$.

Implication 4: Follow-on fund returns
If learning is asymmetric, soft information predicts LP returns in a 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 in the Appendix. The intuition is straightforward. The reason that Implications 2a and 3a hold 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 if so, its size.

Implications 2b, 3b, and 4 potentially allow us to discriminate between symmetric and asymmetric learning and so to test our model. The intuition for Implication 2b is as follows. The GP can only raise a follow-on fund if incumbent LPs have learned that his skill $\mu^i > \mu^*$. 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_1^i:

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

The final return $r^i_{first,final}$ is an increasing function of both the hard information released at $t=1$, H_1^i, and the additional signal H_2^i that becomes public information at $t=2$:

$$r^i_{first,final} = \frac{\frac{1}{2}\left(C_2^i - I_0\right) - M_0}{\frac{1}{2}I_0} = \frac{\frac{1}{2}\left[\left(a+H_1^i+H_2^i\right)\ln(1+I_0) - I_0\right] - M_0}{\frac{1}{2}I_0}.$$

Thus, $r^i_{first,interim}$ fully reveals H_1^i, and given H_1^i, $r^i_{first,final}$ fully reveals H_2^i. In short, both H_1^i and H_2^i (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 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.

1.4 Optimality of asymmetric learning

We end the model by considering the 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.

Asymmetric learning can even lead to more efficient fundraising than symmetric learning, if LPs find it unattractive to invest in the average GP's first fund even at a fund fee of zero:

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

In a risk neutral setting, this would be equivalent to the average NPV of optimally sized first funds being negative: $\max_{I_0} E_{\mu^i}\left(E\left(C_2^i|\mu^i\right)-I_0\right)<0$. Under this condition, no GP would be able to raise a first fund (or any follow-on funds) if learning was symmetric. However, with asymmetric learning, LPs earn informational rents in follow-on funds, which may be sufficient to make up for the expected losses on first funds. Effectively, with asymmetric learning, investment in a first fund gives LPs an option to invest in a follow-on fund. The option value increases in uncertainty about GP skill. If it equals or exceeds the expected loss on first funds (i.e., if there is enough dispersion in GP skill), LPs will invest in first funds despite their negative contribution to expected utility.

The existence of soft information about 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 holdup 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 VC contracts provide exclusive informational rights to incumbent LPs, while prohibiting LPs from sharing such information with outsiders, is consistent with GPs recognizing the benefits of informational holdup. Of course, ex post, GPs who subsequently learn that they have skill have an incentive to signal their type to outside investors when raising a follow-on fund. Signaling does happen in practice, but it is unlikely to have sufficient precision to eliminate the information asymmetry between incumbent and outside investors. For example, "grandstanding," the practice of selling portfolio companies earlier than optimal (Gompers 1996), is unlikely to fully reveal the GP's type as the number of exits 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 grant rights to invest particular amounts at a particular fee should a follow-on fund be raised, suggesting enforcement problems. Courts are also not equipped to determine that the correct GPs raise follow-on funds or that the correct follow-on fund sizes are raised, as this would require courts to gather soft information about GP skill.

2. 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 2,257 U.S. funds raised by 962 VC firms between 1980 and 2002 using data from Thomson Reuters' Venture Economics (VE) and Private Equity Intelligence (PREQIN).[20,21] The annual number of funds raised averages 62 in the 1980s, 106 in the 1990s, and 192 between 2000 and 2002. The average (median) fund raised $111.2 million ($46.0 million) in nominal dollars. Average 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 end of the late 1990s tech boom. Thirty-nine percent of sample funds are first-time funds, and the average fund sequence number is 2.8.[22] We use fund stage focus as a crude control for differences in risk across funds. Fifty-four percent 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 follow-on funds. Because VC funds typically have a ten-year life, we track each sample fund through October 2012, which gives us a minimum of ten 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.[23] This gives a rough estimate of the prevalence of skill in the VC industry of

20 We define all funds listed as focusing on start-up, early-stage, development, late-stage, or expansion investments, and those listed as "venture (general)" or "balanced" funds as VC funds. 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.

21 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.

22 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.

23 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.

Table 2
Descriptive statistics

	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

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 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, as VC funds. In cases in which 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 ten-year) life, net of management and performance fees, using VE and PREQIN data through October 2012.

around a third, assuming GPs go out of business when investors learn that their skill μ^i is below the break-even level, μ^*. 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.

2.1 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 only to GPs and incumbent LPs.

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. VE and PREQIN collect these performance data for dissemination to subscribers, usually in highly aggregated 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.[24] 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 ten years, and our final IRRs are the realized returns after at least ten 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%).[25] There is considerable variation over time in these averages. Whereas 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.

[24] As Ljungqvist and Richardson (2003) show, over a fund's life, performance follows a "J-curve," in the sense that cash-on-cash IRRs in respect of exited investments (rather than reported interim IRRs based on unrealized investments) 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.

[25] The data are thus skewed to the right. However, winsorizing the data does not materially affect our results.

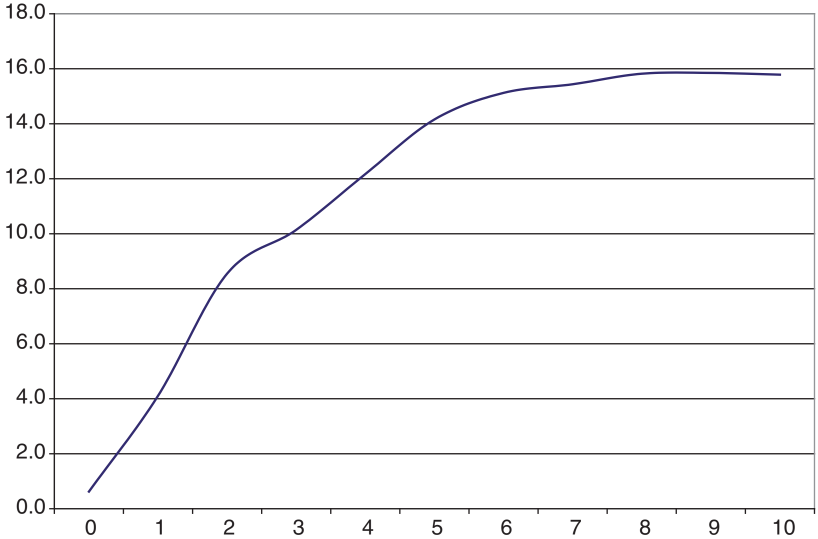

Figure 3
Average interim IRRs over a fund's lifetime
This figure shows the average interim IRR, net of fees, in percent over a fund's ten-year lifetime for a sample of 547 VC funds for which a complete time series of year-by-year interim IRR data is available.

We have interim IRRs for 15,205 fund-years in respect of 944 individual funds.[26] There are frequent 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.[27] There can also be gaps in the middle or toward 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 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 leveling off at a little under 16% in subsequent years.[28]

2.2 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 ten years are up, than do outside

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

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

[28] 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.

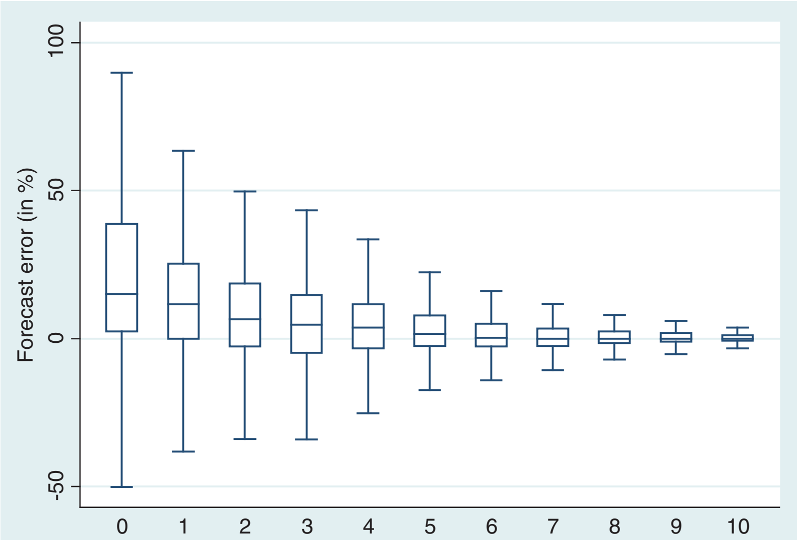

Figure 4
How accurately do interim IRRs forecast final performance?
This figure shows box plots of the distribution of forecast errors (= final IRR – interim IRR, in %) for each year in a fund's life, using all 15,205 fund-years for which interim IRR are available. Each box shows the 75th percentile (the upper hinge of the box), the median (the line drawn inside the box), and the 25th percentile (the lower hinge). The end-points of the whiskers extending from each box denote the 5th and 95th percentiles.

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 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.

2.3 First and follow-on funds

Implications 1 through 4 relate a first fund's interim and final performance to the likelihood that the GP raises a follow-on fund half way through the life of the first fund, and if so, its size and performance. 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 have barely 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 skill. It is thus questionable whether much learning has taken place when GPs raise their second funds; skill could well take more than one fund to be revealed.[29] 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 with any precision. For this reason, our empirical specifications will flexibly distinguish between first and follow-on funds, rather than between first and second funds only.

2.4 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 two years of operation, as mentioned earlier, we require this fund be at least three years old. If the most recent fund is less than three 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 fund 3, we measure the interim IRR of fund 2, if that fund is at least three years old, or else the interim IRR of fund 1. 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 prior-fund interim and final IRRs for 684 follow-on funds. Our performance persistence tests additionally require final IRR data for the follow-on funds themselves.

[29] 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.

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. In practice, conditioning on the performance of older funds does not affect our results.

3. Empirical Analysis

The focus of our empirical analysis is on the role of asymmetric learning and soft information in explaining 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.

3.1 Persistence, learning, and soft information

3.1.1 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 final IRR on log fund size, the previous fund's final IRR, and vintage-year effects.[30] 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.

3.1.2 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

[30] In the model, once the GP's skill is known to the GP and the incumbent LPs, fund size is a sufficient statistic for skill. In practice, fund size may not fully reveal skill for a variety of reasons. We follow standard practice and control for log fund size in our empirical specifications.

Table 3
VC fund performance persistence

	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***
Wald test: ex post IRR = interim IRR	n.a.	n.a.	n.a.	n.a.	13.4***	13.5***
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

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 (2005) results using IRRs and exit rates, respectively. Columns 3 and 4 test Implications 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. n.a. indicates not available.

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. Because VC funds are not traded, traditional risk proxies are unavailable. Instead, we follow Kaplan and Schoar (2005) and include a dummy variable that equals one 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 8.8 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

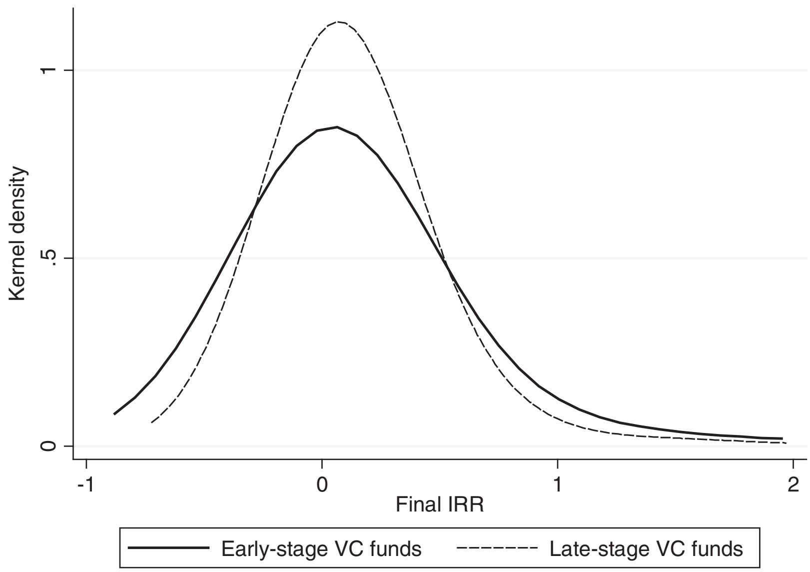

Figure 5
Early- versus late-stage funds
This figure shows the distribution of final IRRs for early- and late-stage VC funds in our sample, respectively. The graph presents a Gaussian kernel density using optimal half-widths and 100 estimation points.

returns significantly predict the future follow-on fund returns. (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.[31] A look at the regression R^2 confirms this. Compared to Column 4, adding ex post returns substantially increases the adjusted R^2, from

31 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.

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 post IRRs than to interim IRRs is the fact (documented in Figures 3 and 4) that interim IRRs tend to rise monotonically over a fund's life 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.[32]

3.2 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 holdup 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.

3.2.1 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

[32] Another potential confound is due to the fact that funds overlap in time and so are subject to similar economic conditions. This affects ex post IRRs, but, by construction, not interim IRRs at fundraising.

Table 4
Effect of learning on fundraising

	Prob(follow-on fund raised)		Log size of follow-on fund	
	(1)	(2)	(3)	(4)
Previous fund's performance				
Interim IRR of fund $N-1$ as	0.270***	0.132***	2.178***	1.785***
of previous year-end	*0.040*	*0.059*	*0.402*	*0.563*
Ex post IRR of fund $N-1$		0.226***		0.450**
		0.067		*0.183*
Controls				
Log fund size	0.195***	0.187***	1.386***	1.417***
	0.025	*0.025*	*0.105*	*0.109*
Diagnostics				
Vintage year FE	n.a.	n.a.	yes	yes
Wald test: all coeff. = 0	125.7***	127.1***	16.9***	14.0***
Wald test: ex post IRR = interim IRR	n.a.	0.7	n.a.	4.1**
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

This table reports tests of Implications 2 and 3 of the model, regarding the effect of performance on future fundraising. In Columns 1 and 2, we estimate a Cox semiparametric 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 three 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 fundraising. It includes all available vintages through 2012; because VC firms have a nonzero 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 is 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, clustered on VC firm, are shown in italics. We use ***, **, and * to denote significance at the 1%, 5%, and 10% level (two-sided), respectively. n.a. indicates not available.

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, roughly doubling the unconditional mean. This supports Implication 2a.

Column 2 additionally conditions on the prior fund's final IRR, which will not be publicly known until, on average, seven 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 ($p=0.001$).[33] The

[33] This echoes the findings of Kaplan and Schoar (2005), who show that ex post IRRs predict fundraising outcomes.

corresponding influence of publicly available interim IRRs, on the other hand, is halved compared to Column 1 ($p=0.024$).

3.2.2 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=0.014$ in Column 4).

3.3 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 publicly 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 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 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 formalizes the informational holdup 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? 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 up the GP, our model applies.

Da Rin and Phalippou's survey of LPs, discussed previously, helps us 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 did not 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 did not re-invest, other investors would be suspicious and would not invest," supporting our informational holdup 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 in which 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.[34] 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 holdup as follows. The practitioner story implies that we should not see persistence in the subset 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.

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

Table 5
Alternative explanation for persistence: Fundraising in good and bad times

	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 Nhas 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***
Wald test: ex post IRR = interim IRR	12.8***	16.1***	11.6***	15.1***
Adjusted R^2	22.7%	23.8%	23.9%	23.6%
No. of observations	302	201	344	308

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.

4. 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 holdup 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 only observe 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 up the GP when he next raises a fund.

Performance is persistent because the holdup problem prevents the GP from raising his fees to the point at which 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 holdup story provides additional supportive evidence for our theory.

Appendix: Derivations and Proofs

Proof of simple example result 1. $E\left(r_2^i|\mu^i\right)=a+\mu^i-\frac{M_2\left(\mu^i\right)}{I}-1$. Thus $\frac{dE\left(r_2^i|\mu^i\right)}{d\mu^i}=1-\frac{M_2'\left(\mu^i\right)}{I}>0$ iff $\frac{M_2'\left(\mu^i\right)}{I}<1$. This is sufficient for $E\left(r_2^i|r_1^i\right)$ to be increasing in r_1^i:

$$E\left(r_2^i|r_1^i\right)=E_{\mu^i}\left(E\left(r_2^i|r_1^i,\mu^i\right)\right)=\int_{-\mu}^{\mu}E\left(r_2^i|\mu^i\right)f\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)d\mu^i$$

which implies

$$\frac{dE\left(r_2^i|r_1^i\right)}{dr_1^i}=\int_{-\mu}^{\mu}E\left(r_2^i|\mu^i\right)\frac{df\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)}{dr_1^i}d\mu^i.$$

Now, $\int_{-\mu}^{\mu}\frac{df\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)}{dr_1^i}d\mu^i=0$ since $f\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)$ is a probability distribution (for any r_1^i). Furthermore, since $\mu^i=r_1^i-\left[a+\varepsilon_1^i-\frac{M_1}{I}-1\right]$ is increasing in r_1^i, there exists a value of μ^i, call it μ^x (which will depend on r_1^i) and for which $\frac{f\left(\mu^i|r_1^i\right)}{dr_1^i}\gtreqless 0$ for $\mu^i\gtreqless\mu^x$. Thus, $\frac{dE\left(r_2^i|r_1^i\right)}{dr_1^i}$ is positive as long as $E\left(r_2^i|\mu^i\right)$ is positive and increasing in μ^i for all values of μ^i, since then the positive values of $\frac{f\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)}{dr_1^i}$ in $\int_{-\mu}^{\mu}E\left(r_2^i|\mu^i\right)\frac{f\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)}{dr_1^i}d\mu^i$ are multiplied by a larger positive number than are the negative values of $\frac{df\left(\mu^i|r_1^i,\mu>\mu^i>\mu^*\right)}{dr_1^i}$.

To formally show that a μ^x with the stated properties exists, note that since $r_1^i|\mu^i=a+\mu^i+\varepsilon_1^i-\frac{M_1}{I}-1\sim N\left(a+\mu^i-\frac{M_1}{I}-1,\sigma^2\right)$,

$$f\left(r_1^i|\mu>\mu^i>-\mu\right)=\int_{-\mu}^{\mu}f\left(r_1^i|\mu^i,\mu>\mu^i>-\mu\right)f\left(\mu^i|\mu>\mu^i>-\mu\right)d\mu^i$$

$$=\int_{-\mu}^{\mu}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}d\mu^i\frac{1}{2\mu}=\frac{1}{2\mu}\left[\Phi\left(z_\mu\right)-\Phi\left(z_{-\mu}\right)\right]$$

and

$$f\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)=f\left(r_1^i|\mu^i,\mu>\mu^i>-\mu\right)\frac{f\left(\mu^i|\mu>\mu^i>-\mu\right)}{f\left(r_1^i|\mu>\mu^i>-\mu\right)}$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}\frac{\frac{1}{2\mu}}{\frac{1}{2\mu}\left[\Phi\left(z_\mu\right)-\Phi\left(z_{-\mu}\right)\right]}=\frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right)-\Phi\left(z_{-\mu}\right)\right]}$$

for $\mu>\mu^i>-\mu$, 0 otherwise, where ϕ and Φ are the pdf and cdf of the standard normal distribution, $z_{\mu^i}=\frac{r_1^i-\left[a+\mu^i-\frac{M_1}{I}-1\right]}{\sigma}$, $z_\mu=\frac{r_1^i-\left[a+\mu-\frac{M_1}{I}-1\right]}{\sigma}$, and $z_{-\mu}=\frac{r_1^i+\left[a-\mu-\frac{M_1}{I}-1\right]}{\sigma}$. Note that this simply says that $\mu^i|r_1^i,\mu>\mu^i>-\mu$ is truncated normal, with truncation at μ and $-\mu$. Since $\phi\left(z_{\mu^i}\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}$, $\frac{d\phi\left(z_{\mu^i}\right)}{dr_1^i}=-\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}\frac{z_{\mu^i}}{\sigma}=-\phi\left(z_{\mu^i}\right)\frac{z_{\mu^i}}{\sigma}$, we have

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

$$=\frac{\frac{1}{\sigma}\phi\left(z_{\mu^i}\right)\frac{1}{\sigma}}{\Phi\left(z_\mu\right)-\Phi\left(z_{-\mu}\right)}\left\{-z_{\mu^i}-\frac{\phi\left(z_\mu\right)-\phi\left(z_{-\mu}\right)}{\Phi\left(z_\mu\right)-\Phi\left(z_{-\mu}\right)}\right\}$$

$$=f\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)\frac{1}{\sigma}\left\{-z_{\mu^i}-\frac{\phi\left(z_\mu\right)-\phi\left(z_{-\mu}\right)}{\Phi\left(z_\mu\right)-\Phi\left(z_{-\mu}\right)}\right\}.$$

The function $f\left(\mu^i|r_1^i,\mu>\mu^i>-\mu\right)\frac{1}{\sigma}$ 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 (as $\frac{\phi(z_\mu)-\phi(z_{-\mu})}{\Phi(z_\mu)-\Phi(z_{-\mu})}$ 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 $\mu^i>\mu^x$.

Proof of Proposition 1. (a) Part (a) is true for any value of p. Consider an offer $\left[(I_{1,split}^{GP}/2,M_{1,split}^{GP}),(I_{1,sole}^{GP},2M_{1,sole}^{GP})\right]$ with fund sizes $I_{1,split}^{GP}$ and $I_{1,sole}^{GP}$ that are different from $I_{1,split}^i$ and $I_{1,sole}^i$. By definition, $I_{1,split}^i$ and $I_{1,sole}^i$ are the joint-surplus-maximizing fund sizes, and so the GP can always make himself better off by changing the proposed fund sizes to $I_{1,split}^i$ and $I_{1,sole}^i$ 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 $(I_{1,split}^i/2,M_1^*\left(\mu^i\right))$ is

$$E\left(U_{split}^{LP}|\mu^i\right)=1-e^{-\gamma W_0^{LP}}E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]+\gamma M_0}|\mu^i\right)$$

$$E\left(e^{-\gamma\frac{1}{2}\left[A_3\ln\left(1+I_{1,split}^i\right)-I_{1,split}^i\right]+\gamma M_1^*\left(\mu^i\right)}|\mu^i\right)$$

$$=1-e^{-\gamma W_0^{LP}}E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]+\gamma M_0}|\mu^i\right)e^{-\gamma\left[b_{split}\left(\mu^i\right)-M_1^*\left(\mu^i\right)\right]}.$$

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

$$E\left(U^{LP}_{sole}|\mu^i\right)=1-e^{-\gamma W^{LP}_0}E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]+\gamma M_0}|\mu^i\right)e^{-\gamma\left[b_{sole}\left(\mu^i\right)-2M^*_1\left(\mu^i\right)\right]}.$$

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

$$1-e^{-\gamma\left[W^{GP}_0+2M_0+2M^{LP,*}_{1,split}\right]}=p\left[1-e^{-\gamma\left[W^{GP}_0+2M_0\right]}\right]+(1-p)\left[1-e^{-\gamma\left[W^{GP}_0+2M_0+2M^{GP,*}_{1,split}\right]}\right]\Longleftrightarrow$$

$$e^{-\gamma 2M^{LP,*}_{1,split}}=p+(1-p)e^{-\gamma 2M^{GP,*}_{1,split}}\Longleftrightarrow$$

$$e^{\gamma 2M^{GP,*}_{1,split}}=\frac{1-p}{e^{-\gamma 2M^{LP,*}_{1,split}}-p}$$

Each LP's indifference condition is

$$1-e^{-\gamma W^{LP}_0}E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]+\gamma M_0}|\mu^i\right)E\left(e^{-\gamma\left[\frac{1}{2}\left(A_3\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)-M^{GP,*}_{1,split}\right]}|\mu^i\right)$$

$$=p\left[1-e^{-\gamma W^{LP}_0}E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]+\gamma M_0}|\mu^i\right)\right]$$

$$+(1-p)\left[\begin{array}{c}1-e^{-\gamma W^{LP}_0}E\left(e^{-\gamma\frac{1}{2}[A_2\ln(1+I_0)-I_0]+\gamma M_0}|\mu^i\right)\\ E\left(e^{-\gamma\left[\frac{1}{2}\left(A_3\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)-M^{LP,*}_{1,split}\right]}|\mu^i\right)\end{array}\right]$$

$\Longleftrightarrow$

$$E\left(e^{-\gamma\left[\frac{1}{2}\left(A_3\ln\left(1+I^i_{1,split}\right)-I^i_{1,split}\right)-M^{GP,*}_{1,split}\right]}|\mu^i\right)$$

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

$\Longleftrightarrow$

$$e^{-\gamma b_{split}\left(\mu^i\right)+\gamma M^{GP,*}_{1,split}}=p+(1-p)e^{-\gamma b_{split}\left(\mu^i\right)+\gamma M^{LP,*}_{1,split}}.$$

Combining the two indifference conditions implies

$$e^{-\gamma b_{split}\left(\mu^i\right)}\left(\frac{1-p}{e^{-\gamma 2M^{LP,*}_{1,split}}-p}\right)^{1/2}=p+(1-p)e^{-\gamma b_{split}\left(\mu^i\right)+\gamma M^{LP,*}_{1,split}}.$$

Denote $e^{-\gamma M^{LP,*}_{1,split}}$ by x and $e^{\gamma b_{split}\left(\mu^i\right)}$ by y. Then the above can be rewritten as

$$\frac{1}{y}\left(\frac{1-p}{x^2-p}\right)^{1/2}=p+(1-p)\frac{1}{y}\frac{1}{x}\Longleftrightarrow$$

$$0=py^2x^4+2y(1-p)x^3+\left(-1+p-p^2y^2\right)x^2-2py(1-p)x-(1-p)^2.$$

This is a continuous function of p. Thus, as p goes to zero, x solves $0=2yx^3-x^2-1$ which has one real solution:

$$x=-\frac{-1}{6y}-\frac{1}{6y}\left(\frac{1}{2}\left[-2-108y^2+\sqrt{\left[-2-108y^2\right]^2-4}\right]\right)^{1/3}$$
$$-\frac{1}{6y}\left(\frac{1}{2}\left[-2-108y^2-\sqrt{\left[-2-108y^2\right]^2-4}\right]\right)^{1/3}.$$

which will be a function of μ^i since $y=e^{\gamma b_{split}\left(\mu^i\right)}$.

Given this solution for x, $M_{1,split}^{LP,*}=\frac{-\ln\left(x\left(\mu^i\right)\right)}{\gamma}$ and $M_{1,split}^{GP,*}$ equals $M_{1,split}^{LP,*}$ 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 $M_1^*\left(\mu^i\right)$. Expressing $M_1^*\left(\mu^i\right)$ as a fraction $g\left(\mu^i\right)$ of the LP's (pre-fee) risk adjusted expected cash flow, $M_1^*\left(\mu^i\right)=g\left(\mu^i\right)b_{split}\left(\mu^i\right)$, we have that $g\left(\mu^i\right)=-\frac{\ln(x)}{\gamma b_{split}\left(\mu^i\right)}=-\frac{\ln(x)}{\ln(y)}$.

One can verify numerically that (for any value of γ) $g\left(\mu^i\right)$ is monotonically decreasing in μ^i, $g\left(\mu^i\right)\to 1/2$ as $b_{split}\left(\mu^i\right)\to 0$, and $g\left(\mu^i\right)\to 1/3$ as $b_{split}\left(\mu^i\right)\to\infty$. To derive the limits, first note that (by x solving $0=2yx^3-x^2-1$), as $b_{split}\left(\mu^i\right)$ goes to zero, y goes to one, and thus x goes to one. Furthermore, as $b_{split}\left(\mu^i\right)$ goes to infinity, y goes to infinity, and thus x goes to 0. Second, rewrite the equation $0=2yx^3-x^2-1$ to express y as a function of x, $y=\frac{1+x^2}{2x^3}$, which implies $g\left(\mu^i\right)=-\frac{\ln(x)}{\ln(y)}=-\frac{\ln(x)}{\ln\left(1+x^2\right)-\ln\left(2x^3\right)}$. Thus, using l'Hopital's rule (both the numerator and denominator goes to zero for $x\to 1$, while they both go to $-\infty$ for $x\to 0$) we get

$$g\left(\mu^i\right)=-\frac{\ln(x)}{\ln(y)}=-\frac{\ln(x)}{\ln\left(1+x^2\right)-\ln\left(2x^3\right)}\to-\frac{\frac{1}{x}}{\frac{2x}{1+x^2}-\frac{6x^2}{2x^3}}$$
$$=-\frac{1}{\frac{2x^2}{1+x^2}-3}=\begin{cases}\frac{1}{2} & \text{for } x\to 1\ (b_{split}\left(\mu^i\right)\to 0)\\ \frac{1}{3} & \text{or } x\to 0\ (b_{split}\left(\mu^i\right)\to\infty)\end{cases}$$

(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 $M_{1,split}^{GP}$ above $M_1^*\left(\mu^i\right)$, or $M_{1,sole}^{GP}$ above $2M_1^*\left(\mu^i\right)$, since LPs will reject all such offers. Furthermore, under the proposed strategies, LPs accept the GP's split offer $\left[I_{1,split}^i/2, M_1^*\left(\mu^i\right)\right]$, 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 $M_{1,split}^{LP_a}$ ($M_{1,sole}^{LP_a}$) below $M_1^*\left(\mu^i\right)$ ($2M_1^*\left(\mu^i\right)$) 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 one LP's sole offer offered $2M_1^*\left(\mu^i\right)-\varepsilon$, the other LP could reduce his sole offer in the same way and adjust his offered split fee to $M_1^*\left(\mu^i\right)-(1-\delta)\varepsilon$, for δ arbitrarily small, and still be asked to invest as part of a split outcome. But the first LP would then also adjust his split offer to this value, with the result that 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 $M_{1,split}^{LP_a}$ above M_1^*. If LP$_a$ did so, then the GP would accept both LPs' split offers (earning fees of $M_{1,split}^{LP_a}+M_1^*>2M_1^*$), and therefore LP$_a$ would end up with the same investment of $I_{1,split}^i/2$ but would pay a higher fee. In addition, LP$_a$ has no incentive to increase

$M^{LPa}_{1,sole}$ 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 $(I^i_{1,sole}, M^{LPa}_{1,sole})$ with $M^{LPa}_{1,sole} > 2M^*_1$.

Proof of Corollary 2. This follows directly from the fact that $g\left(\mu^i\right)$ is less than one and decreasing in μ^i and that $b\left(\mu^i\right)$ is positive for $\mu^i > \mu^*$ and increasing in μ^i ($I^i_{1,split}$ maximizes $b_{split}\left(\mu^i\right)$ so by the envelope theorem $\frac{db_{split}\left(\mu^i\right)}{dE\left(A_3|\mu^i\right)} = \frac{\partial b_{split}\left(\mu^i\right)}{\partial E\left(A_3|\mu^i\right)} = \frac{1}{2}\ln\left(1+I^i_{1,split}\right) > 0$ and $E\left(A_3|\mu^i\right) = a+2\mu^i$).

Proof of Implication 1. We prove Implication 1b. Since the risk-adjustment $\frac{1}{4}\gamma\sigma^2 I^i_{1,split}$ is increasing in μ^i, Implication 1a follows immediately from Implication 1b. Proving Implication 1b requires us to prove that the expectation of

$$r^i_{follow-on,final,risk-adj} = \frac{\frac{1}{2}\left(C^i_3 - I^i_{1,split}\right) - M^*_1\left(\mu^i\right) - \frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2}{\frac{1}{2}I^i_{1,split}}$$

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. Using the expression $b_{split}\left(\mu^i\right)$ from Proposition 1,

$$E\left(r^i_{follow-on,final,risk-adj}|\mu^i, \mu > \mu^i > \mu^*\right) = \frac{b_{split}\left(\mu^i\right) - M^*_1\left(\mu^i\right)}{\frac{1}{2}I^i_{1,split}} = \frac{\left(1-g\left(\mu^i\right)\right)b_{split}\left(\mu^i\right)}{\frac{1}{2}I^i_{1,split}}.$$

From Corollary 2, $\left(1-g\left(\mu^i\right)\right)b_{split}\left(\mu^i\right)$ is positive and increasing in μ^i. Thus, $E\left(r^i_{follow-on,final,risk-adj}|\mu^i, \mu > \mu^i > \mu^*\right)$ is positive. Since $g\left(\mu^i\right)$ is decreasing in μ^i (from Proposition 1), 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 $\frac{b_{split}\left(\mu^i\right)}{I^i_{1,split}}$ is increasing in μ^i. Using

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

and

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

we get

$$\frac{d}{d\mu^i}\left(\frac{b_{split}\left(\mu^i\right)}{I^i_{1,split}}\right) = \frac{\partial\left(\frac{b_{split}\left(\mu^i\right)}{I^i_{1,split}}\right)}{\partial I_1\left(\mu^i\right)}\frac{dI^i_{1,split}}{d\mu^i} + \frac{\partial\left(\frac{b_{split}\left(\mu^i\right)}{I^i_{1,split}}\right)}{\partial E\left(A_3|\mu^i\right)}\frac{dE\left(A_3|\mu^i\right)}{d\mu^i}$$

$$= \frac{-b_{split}\left(\mu^i\right)}{\left(I^i_{1,split}\right)^2}\frac{dI^i_{1,split}}{d\mu^i} + \frac{1}{2}\frac{\ln\left(1+I^i_{1,split}\right)}{I^i_{1,split}}\frac{dE\left(A_3|\mu^i\right)}{d\mu^i}$$

since $\frac{db_{split}\left(\mu^i\right)}{dI^i_{1,split}} = 0$ by $I^i_{1,split}$ maximizing $b_{split}\left(\mu^i\right)$. Furthermore, from the expression for $I^i_{1,split}$,

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

so

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

Therefore,

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

which is true given the expression for $b_{split}\left(\mu^i\right)$ since $I^i_{1,split} > \ln\left(1+I^i_{1,split}\right)$ for any $I^i_{1,split} > 0$ and $-\frac{1}{8}\gamma\sigma^2\left(I^i_{1,split}\right)^2 < 0$.

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 $r^i_{first,interim}$.

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

Note that

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

since $H^i_2 = \mu^i+\upsilon^i = H^i_1-\varepsilon^i+\upsilon^i$. Therefore, $E\left(r^i_{follow-on,final,risk-adj}|r^i_{first,interim},\mu > \mu^i > \mu^*\right)$ will be increasing in $r^i_{first,interim}$ iff $E\left(r^i_{follow-on,final,risk-adj}|H^i_1,\mu^i > \mu^*\right)$ is increasing in H^i_1. So we are interested in

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

and

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

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

Now, $\int_{\mu^*}^{\mu}\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}d\mu^i=0$ since $f\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)$ is a probability distribution (for any H_1^i). Furthermore, since $\mu^i=H_1^i-\varepsilon^i$ is increasing in H_1^i, there exists a value of μ^i, call it μ^x (which will depend on H_1^i) and for which $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}\gtreqless 0$ for $\mu^i\gtreqless\mu^x$ (we omit the formal derivation of μ^x since it is very similar to the derivation of μ^x in the proof of Result 1 for the simple example). Thus, $\frac{d}{dH_1^i}E\left(r^i_{follow-on,final,risk-adj}|H_1^i,\mu>\mu^i>\mu^*\right)$ is positive as long as $E\left(r^i_{follow-on,final,risk-adj}|\mu^i,\mu>\mu^i>\mu^*\right)$ is positive and increasing in μ^i for all values of μ^i (which is true by Step 1), since then the positive values of $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}$ in $\int_{\mu^*}^{\mu}E\left(r^i_{follow-on,final,risk-adj}|\mu^i\right)\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>\mu^*\right)}{dH_1^i}d\mu^i$ 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 from the proof of Implication 1, step 2, that

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

This implies that $r^i_{first,interim}$ (along with I_0 and M_0) fully reveals H_1^i. Since H_1^i 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 $P\left(\mu^i>\mu^*|H_1^i,\mu>\mu^i>-\mu\right)$ is increasing in H_1^i.

Since $P\left(\mu^i>\mu^*|H_1^i,\mu>\mu^i>-\mu\right)=\int_{\mu^*}^{\mu}f\left(\mu^i|H_1^i,\mu>\mu^i>-\mu\right)d\mu^i$ we have

$$\frac{dP\left(\mu^i>\mu^*|H_1^i,\mu>\mu^i>-\mu\right)}{dH_1^i}=\int_{\mu^*}^{\mu}\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>-\mu\right)}{dH_1^i}d\mu^i.$$

Now, $\int_{-\mu}^{\mu}\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>-\mu\right)}{dH_1^i}d\mu^i=0$ since $f\left(\mu^i|H_1^i,\mu>\mu^i>-\mu\right)$ is a probability distribution (for any H_1^i). Furthermore, since $\mu^i=H_1^i-\varepsilon^i$ is increasing in H_1^i, there exists a value of μ^i, call it μ^x (which will depend on H_1^i) and for which $\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>-\mu\right)}{dH_1^i}\gtreqless 0$ for $\mu^i\gtreqless\mu^x$ (we omit the formal derivation of μ^x since it is very similar to the derivation of μ^x in the proof of Result 1 for the simple example). Thus, $\int_{\mu^*}^{\mu}\frac{df\left(\mu^i|H_1^i,\mu>\mu^i>-\mu\right)}{dH_1^i}d\mu^i$ is positive (as long as $\mu^*>-\mu$).

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.

Without informational holdup, 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:

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

A given GP thus sets $M_1\left(H_1^i\right)=E\left(A_3^i|H_1^i\right)\ln\left(1+I_1\left(H_1^i\right)\right)-I_1\left(H_1^i\right)$. 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

$$E\left(A_3^i|H_1^i\right)=E\left(a+H_{2,follow-on}^i+H_{3,follow-on}^i|H_1^i\right)=a+2E\left(\mu^i|H_1^i\right)=a+2H_1^i.$$

Maximizing the NPV of the fund thus results in:

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

$$M_1\left(H_1^i\right)^{\text{Sym info}} = \left(a+2H_1^i\right)\ln\left(a+2H_1^i\right)-\left[a+2H_1^i-1\right] \text{ for } H_1^i>\frac{1-a}{2}, \text{ zero otherwise.}$$

The outcome for first funds is similar. At $t=0$, LPs' participation constraint is $a\ln(1+I_0)-I_0-M_0=0$. A given GP thus sets $M_0=a\ln(1+I_0)-I_0$. 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

$$I_0^{\text{Sym info}}=a-1,\ M_0^{\text{Sym info}}=a\ln(a)-[a-1].$$

Thus,

$$r_{first,interim}^i=\frac{E(C_2^i|H_1^i)-I_0^{\text{Sym info}}-M_0^{\text{Sym info}}}{I_0^{\text{Sym info}}}=\frac{\left(a+2H_1^i\right)\ln(a)-a\ln(a)}{a-1}=\frac{2H_1^i\ln(a)}{a-1}.$$

This implies that $r_{first,interim}^i$ fully reveals H_1^i. Since follow-on funds are raised iff $H_1^i>\frac{1-a}{2}$, this means that they are raised iff

$$r_{first,interim}^i>\frac{(1-a)\ln(a)}{a-1}=-\ln(a).$$

The right-hand-side expression is a constant that is known at $t=0$. Denote it by $r_{first,interim}^{i,*}$. Thus, $P\left(H_1^i>\frac{1-a}{2}|r_{first,interim}^i\right)=0$ for $r_{first,interim}^i\leq r_{first,interim}^{i,*}$ and $P\left(H_1^i>\frac{1-a}{2}|r_{first,interim}^i\right)=1$ for $r_{first,interim}^i>r_{first,interim}^{i,*}$, 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, $r_{first,interim}^i$.

(b) Start from our assumptions that

$$C_2^i=A_2^i\ln\left(1+I_0^i\right)$$

$$A_2^i=a+H_1^i+H_2^i=a+2\mu^i+\varepsilon^i+v^i$$

$$H_1^i=\mu^i+\varepsilon^i,\ H_2^i=\mu^i+v^i.$$

Note that

$$r^i_{first,final} = \frac{\frac{1}{2}\left(C^i_2 - I_0\right) - M_0}{\frac{1}{2}I_0}$$

$$= \frac{\frac{1}{2}\left(\left(a + H^i_1 + H^i_2\right)\ln(1+I_0) - I_0\right) - M_0}{\frac{1}{2}I_0}$$

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

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. Since H^i_2 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 $P\left(\mu^i > \mu^* | H^i_1, H^i_2, \mu > \mu^i > -\mu\right)$ is increasing in H^i_2.

Since $P\left(\mu^i > \mu^* | H^i_1, H^i_2, \mu > \mu^i > -\mu\right) = \int_{\mu^*}^{\mu} f\left(\mu^i | H^i_1, H^i_2, \mu > \mu^i > -\mu\right) d\mu^i$ we have

$$\frac{dP\left(\mu^i > \mu^* | H^i_1, H^i_2, \mu > \mu^i > -\mu\right)}{dH^i_2} = \int_{\mu^*}^{\mu} \frac{df\left(\mu^i | H^i_1, H^i_2, \mu > \mu^i > -\mu\right)}{dH^i_2} d\mu^i.$$

Now, $\int_{-\mu}^{\mu} \frac{df\left(\mu^i | H^i_1, H^i_2, \mu > \mu^i > -\mu\right)}{dH^i_2} d\mu^i = 0$ since $f\left(\mu^i | H^i_1, H^i_2, \mu > \mu^i > -\mu\right)$ is a probability distribution (for any H^i_1, H^i_2). Furthermore, since $\mu^i = \frac{1}{2}\left(H^i_1 + H^i_2\right) - \frac{1}{2}\left(\varepsilon^i + v^i\right)$ is increasing in H^i_2, there exists a value of μ^i, call it μ^x (which will depend on H^i_1 and H^i_2) and for which $\frac{df\left(\mu^i | H^i_1, H^i_2, \mu > \mu^i > -\mu\right)}{dH^i_2} \gtreqless 0$ for $\mu^i \gtreqless \mu^x$. Thus, $\int_{\mu^*}^{\mu} \frac{df\left(\mu^i | H^i_1, \mu > \mu^i > -\mu\right)}{dH^i_1} d\mu^i$ is positive (as long as $\mu^* > -\mu$).

To formally show that a μ^x with the stated properties exists, note that the information about μ^i in H^i_1 and H^i_2 can be summarized by the average $H^i = \frac{1}{2}\left(H^i_1 + H^i_2\right) = \mu^i + \frac{1}{2}\left(\varepsilon^i + v^i\right)$, which implies $H^i | \mu^i \sim N\left(\mu^i, \sigma^2_H\right)$ with $\sigma^2_H = \frac{1}{4}\left(\sigma^2_\varepsilon + \sigma^2_v\right)$. Therefore,

$$f\left(H^i | \mu > \mu^i > -\mu\right) = \int_{-\mu}^{\mu} f\left(H^i | \mu^i, \mu > \mu^i > -\mu\right) f\left(\mu^i | \mu > \mu^i > -\mu\right) d\mu^i$$

$$= \int_{-\mu}^{\mu} \frac{1}{\sqrt{2\pi\sigma^2_H}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} d\mu^i \frac{1}{2\mu} = \frac{1}{2\mu}\left[\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)\right]$$

with $z_{\mu^i} = \frac{H^i - \mu^i}{\sigma_H}$, $z_\mu = \frac{H^i - \mu}{\sigma_H}$, and $z_{-\mu} = \frac{H^i + \mu}{\sigma_H}$. Thus,

$$f\left(\mu^i | H^i, \mu > \mu^i > -\mu\right) = f\left(H^i | \mu^i, \mu > \mu^i > -\mu\right) \frac{f\left(\mu^i | \mu > \mu^i > -\mu\right)}{f\left(H^i | \mu > \mu^i > -\mu\right)}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2_H}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} \frac{\frac{1}{2\mu}}{\frac{1}{2\mu}\left[\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)\right]} = \frac{\frac{1}{\sigma_H}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)\right]}$$

for $\mu > \mu^i > -\mu$, and 0 otherwise. Thus, $\mu^i | H^i_1, H^i_2, \mu > \mu^i > -\mu$ has a truncated normal distribution with truncation at $-\mu$ and μ. Note that $\phi\left(z_{\mu^i}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2}$ implies

$\frac{d\phi\left(z_{\mu^i}\right)}{dH_2^i} = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(z_{\mu^i}\right)^2} \frac{z_{\mu^i}}{2\sigma_H} = -\phi\left(z_{\mu^i}\right)\frac{z_{\mu^i}}{2\sigma_H}$. Therefore,

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

$$- \frac{\frac{1}{\sigma_H}\phi\left(z_{\mu^i}\right)}{\left[\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)\right]^2}\left[\phi\left(z_\mu\right)\left(\frac{1}{2\sigma_H}\right) - \phi\left(z_{-\mu}\right)\left(\frac{1}{2\sigma_H}\right)\right]$$

$$= \frac{\frac{1}{\sigma_H}\phi\left(z_{\mu^i}\right)\frac{1}{2\sigma_H}}{\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)}\left\{-z_{\mu^i} - \frac{\phi\left(z_\mu\right) - \phi\left(z_{-\mu}\right)}{\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)}\right\}$$

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

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 $\left\{-z_{\mu^i} - \frac{\phi\left(z_\mu\right) - \phi\left(z_{-\mu}\right)}{\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)}\right\}$ is increasing in μ^i (since $\frac{\phi\left(z_\mu\right) - \phi\left(z_{-\mu}\right)}{\Phi\left(z_\mu\right) - \Phi\left(z_{-\mu}\right)}$ 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 $\frac{df\left(\mu^i|H_1^i, H_2^i, \mu > \mu^i > -\mu\right)}{dH_2^i} \gtrless 0$ for $\mu^i \gtrless \mu^x$.

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 $E\left(A_3^i|\mu^i\right) > 1$ for $\mu^i > \mu^*$. Since $A_3^i = a + H_{2,follow-on}^i + H_{3,follow-on}^i = a + 2\mu^i + \varepsilon_{follow-on}^i + \upsilon_{follow-on}^i$, we have $E\left(A_3^i|\mu^i\right) = a + 2\mu^i$, so $I_{1,split}^i$ is positive and increasing in μ^i.

From the proof of Implication 1, step 2, $r_{first,interim}^i$ is given by

$$r_{first,interim}^i = \frac{\frac{1}{2}\left(\left(a + 2H_1^i\right)\ln(1 + I_0) - I_0\right) - M_0}{\frac{1}{2}I_0}.$$

so $r_{first,interim}^i$ (along with I_0 and M_0) fully reveals H_1^i and H_1^i is positively related to $r_{first,interim}^i$. It follows that $E\left(I_{1,split}^i|r_{first,interim}^i, \mu > \mu^i > -\mu^*\right)$ is increasing in $r_{first,interim}^i$ iff $E\left(I_{1,split}^i|H_1^i, \mu > \mu^i > -\mu^*\right)$ is increasing in H_1^i. Furthermore,

$$E\left(I_{1,split}^i|H_1^i, \mu > \mu^i > -\mu^*\right) = \int_{\mu^*}^{\mu} I_{1,split}^i f\left(\mu^i|H_1^i, \mu > \mu^i > \mu^*\right) d\mu^i$$

$$\frac{dE\left(I_{1,split}^i|H_1^i, \mu > \mu^i > -\mu^*\right)}{dH_1^i} = \int_{\mu^*}^{\mu} I_{1,split}^i \frac{df\left(\mu^i|H_1^i, \mu > \mu^i > \mu^*\right)}{dH_1^i} d\mu^i$$

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

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

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

Combining these two expressions, we get

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

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

(b) If raised, the follow-on fund's size is $I^i_{1,split}$ which from the proof of Implication 3a, the asymmetric learning case, is positive and increasing in μ^i.

From the proof of Implication 2b, $r^i_{first,interim}$ fully reveals H_1^i and given H_1^i, $r^i_{first,final}$ fully reveals H_2^i. Since H_2^i is positively related to $r^i_{first,final}$, it follows that $E\left(I^i_{1,split}|r^i_{first,interim},r^i_{first,final},\mu>\mu^i>-\mu^*\right)$ is increasing in $r^i_{first,final}$ iff $E\left(I^i_{1,split}|H_1^i,H_2^i,\mu>\mu^i>-\mu^*\right)$ is increasing in H_2^i. Furthermore,

$$E\left(I^i_{1,split}|H_1^i,H_2^i,\mu>\mu^i>-\mu^*\right)=\int_{\mu^*}^{\mu}I^i_{1,split}f\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>\mu^*\right)d\mu^i$$

$$\frac{dE\left(I^i_{1,split}|H_1^i,H_2^i,\mu>\mu^i>-\mu^*\right)}{dH_2^i}=\int_{\mu^*}^{\mu}I^i_{1,split}\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>\mu^*\right)}{dH_2^i}d\mu^i.$$

Now, $\int_{\mu^*}^{\mu}\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>\mu^*\right)}{dH_2^i}d\mu^i=0$ since $f\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>\mu^*\right)$ is a probability distribution (for any H_1^i, H_2^i). Furthermore, since $\mu^i=\frac{1}{2}\left(H_1^i+H_2^i\right)-\frac{1}{2}\left(\varepsilon^i+v^i\right)$ is increasing in H_2^i, there exists a value of μ^i, call it μ^x (which will depend on H_1^i and H_2^i) and for which $\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>\mu^*\right)}{dH_2^i}\gtreqless 0$ for $\mu^i\gtreqless\mu^x$ (we omit the formal derivation of μ^x since it is very similar to the derivation of μ^x in the proof of Implication 2b. Thus, $\frac{dE\left(I^i_{1,split}|H_1^i,H_2^i,\mu>\mu^i>-\mu^*\right)}{dH_2^i}$ is positive since $I^i_{1,split}$ is positive and increasing in μ^i for all $\mu^i>\mu^*$, 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 $\int_{\mu^*}^{\mu}I^i_{1,split}\frac{df\left(\mu^i|H_1^i,H_2^i,\mu>\mu^i>\mu^*\right)}{dH_2^i}d\mu^i$ 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

$$1+r^{i}_{follow-on,final}=\frac{\frac{1}{2}\left(C_{3}^{i}-I_{1,split}^{i}\right)-M_{1}^{*}\left(\mu^{i}\right)}{\frac{1}{2}I_{1,split}^{i}}.$$

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 $\frac{1}{4}\gamma\sigma^{2}I_{1,split}^{i}$ is positive and increasing in μ^{i}.

From the proof of Implication 2b, $r^{i}_{first,interim}$ fully reveals H_{1}^{i} and given H_{1}^{i}, $r^{i}_{first,final}$ fully reveals H_{2}^{i}. Since H_{2}^{i} is positively related to $r^{i}_{first,final}$, 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 $E\left(r^{i}_{follow-on,final,risk-adj}|H_{1}^{i},H_{2}^{i},\mu>\mu^{i}>-\mu^{*}\right)$ is increasing in H_{2}^{i}. Furthermore,

$$\begin{aligned}&E\left(r^{i}_{follow-on,final,risk-adj}|H_{1}^{i},H_{2}^{i},\mu>\mu^{i}>-\mu^{*}\right)\\ =&\int_{\mu^{*}}^{\mu}E\left(r^{i}_{follow-on,final,risk-adj}|\mu^{i},\mu>\mu^{i}>\mu^{*}\right)f\left(\mu^{i}|H_{1}^{i},H_{2}^{i},\mu>\mu^{i}>\mu^{*}\right)d\mu^{i}\end{aligned}$$

so

$$\begin{aligned}&\frac{dE\left(r^{i}_{follow-on,final,risk-adj}|H_{1}^{i},H_{2}^{i},\mu>\mu^{i}>-\mu^{*}\right)}{dH_{2}^{i}}\\ =&\int_{\mu^{*}}^{\mu}E\left(r^{i}_{follow-on,final,risk-adj}|\mu^{i},\mu>\mu^{i}>\mu^{*}\right)\frac{df\left(\mu^{i}|H_{1}^{i},H_{2}^{i},\mu>\mu^{i}>\mu^{*}\right)}{dH_{2}^{i}}d\mu^{i}\end{aligned}$$

This is positive following the same steps as in the proof of Implication 3b (simply replace $I_{1,split}^{i}$ by $E\left(r^{i}_{follow-on,final,risk-adj}|\mu^{i},\mu>\mu^{i}>\mu^{*}\right)$).

Nash bargaining in our main model, with one GP and two LPs per fund:

Consider what the fund fee in follow-on funds would be under symmetric Nash bargaining between one GP and two LPs. Since it is in neither party's interest to deviate from this, we assume that the follow-on fund size is set to maximize the joint payoff as above. The bargaining is thus over the follow-on fund fee paid by each LP, $M_{1}\left(\mu^{i}\right)^{Nash}$.

With symmetric Nash bargaining the objective to be maximized at the start of the follow-on fund is:

$$\begin{aligned}&\left[E\left(U^{GP}|\mu^{i},\varepsilon^{i}\right)-E\left(U^{GP}|\mu^{i},\varepsilon^{i}\right)^{\text{No fund}}\right]\left[E\left(U^{LP}|\mu^{i},\varepsilon^{i}\right)-E\left(U^{LP}|\mu^{i},\varepsilon^{i}\right)^{\text{No fund}}\right]^{2}\\ =&\left[1-e^{-\gamma\left[W_{0}^{GP}+2M_{0}+2M_{1}\right]}-\left(1-e^{-\gamma\left[W_{0}^{GP}+2M_{0}\right]}\right)\right]\\ &\times\left[\begin{array}{c}1-e^{-\gamma\left[W_{0}^{LP}-M_{0}-M_{1}\right]}E\left(e^{-\gamma\frac{1}{2}\left[A_{2}\ln(1+I_{0})-I_{0}\right]}|\mu^{i},\varepsilon^{i}\right)E\left(e^{-\gamma\frac{1}{2}\left[A_{3}\ln(1+I_{1})-I_{1}\right]}|\mu^{i},\varepsilon^{i}\right)\\ -\left(1-e^{-\gamma\left[W_{0}^{LP}-M_{0}\right]}E\left(e^{-\gamma\frac{1}{2}\left[A_{2}\ln(1+I_{0})-I_{0}\right]}|\mu^{i},\varepsilon^{i}\right)\right)\end{array}\right]^{2}\\ =&e^{-\gamma\left[W_{0}^{GP}+2M_{0}\right]}\left[e^{-\gamma\left[W_{0}^{LP}-M_{0}\right]}E\left(e^{-\gamma\frac{1}{2}\left[A_{2}\ln(1+I_{0})-I_{0}\right]}|\mu^{i},\varepsilon^{i}\right)\right]^{2}\\ &\times\left(1-e^{-\gamma 2M_{1}}\right)\left[1-e^{\gamma M_{1}}E\left(e^{-\gamma\frac{1}{2}\left[A_{3}\ln(1+I_{1})-I_{1}\right]}|\mu^{i},\varepsilon^{i}\right)\right]^{2}\end{aligned}$$

where "No fund" refers to the outcome if the follow-on fund is not raised and where the second term is squared since there are two LPs. Maximizing this is equivalent to maximizing

$$\left(1-e^{-\gamma 2M_1}\right)\left(1-e^{\gamma M_1}e^{-\gamma b_{split}\left(\mu^i\right)}\right)^2$$

where $b_{split}\left(\mu^i\right)$ is defined as in Proposition 1. The first-order condition for M_1 is

$$\gamma 2e^{-\gamma 2M_1}\left(1-e^{\gamma M_1}e^{-\gamma b_{split}\left(\mu^i\right)}\right)^2+\left(1-e^{-\gamma 2M_1}\right)2\left(1-e^{\gamma M_1}e^{-\gamma b_{split}\left(\mu^i\right)}\right)\left(-\gamma e^{\gamma M_1}e^{-\gamma b_{split}\left(\mu^i\right)}\right)=0\Longleftrightarrow$$

$$e^{-\gamma 2M_1}\left(1-e^{\gamma M_1}e^{-\gamma b_{split}\left(\mu^i\right)}\right)+\left(1-e^{-\gamma 2M_1}\right)\left(-e^{\gamma M_1}e^{-\gamma b_{split}\left(\mu^i\right)}\right)=0\Longleftrightarrow$$

$$M_1\left(\mu^i\right)^{Nash}=\frac{1}{3}b\left(\mu^i\right).$$

References

Anton, J., and D. Yao. 1989. Split awards, procurement, and innovation. *RAND Journal of Economics* 20:538–52.

Berk, J., and R. Green. 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy*

Binmore, K., A. Rubinstein, and A. Wolinsky. 1986. The Nash bargaining solution in economic modelling. *RAND Journal of Economics* 17:176–88.

Binmore, K., M. J. Osborne, and A. Rubinstein. 1992. Non-cooperative models of bargaining. In *Handbook of game theory with economic applications, Vol. 1*. Eds. R. J. Aumann and S. Hart. Amsterdam, The Netherlands: North Holland.

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. 2012. Pay for performance from future fund flows: The case of private equity. *Review of Financial Studies* 25:3259–304.

Cornelli, F., Z. Kominek, and A. Ljungqvist. 2013. Monitoring managers: Does it matter? *Journal of Finance* 68:431–81.

Da Rin, M., and L. Phalippou. 2011. On the heterogeneity of private equity investors. 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* 43:133–56.

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. Working Paper, Columbia University.

Kaplan, S. N., and A. Schoar. 2005. Private equity performance: Returns, persistence, and capital flows. *Journal of Finance* 61:1791–823.

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* 62:731–64.

Ljungqvist, A., and M. Richardson. 2003. The cash flow, return, and risk characteristics of private equity. Working Paper, NBER No. 9454.

Ljungqvist, A., M. Richardson, and D. Wolfenzon. 2007. The investment behavior of buyout fund managers. Working Paper, New York University.

Malkiel, B. G. 1995. Returns from investing in equity mutual funds 1971-1991. *Journal of Finance* 50:549–72.

Rajan, R. 1992. Insiders and outsiders: The choice between informed and arm's length debt. *Journal of Finance* 47:1367–400.

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* 45:1069–87.

Sorensen, M., N. Wang, and J. Yang. 2012. Valuing private equity. Working Paper, Columbia University.