A Theory of Liquidity in Private Equity*

Vincent Maurin† David T. Robinson‡ Per Strömberg§

This draft: April, 2020

Abstract

We develop a model of private equity in which many empirical patterns arise endogenously. Our model rests solely on two critical features of this market: moral hazard for General partners (GPs) and illiquidity risk for Limited Partners (LPs). The equilibrium fund structure incentivizes GPs with a share in the fund and compensates LPs with an illiquidity premium. GPs may inefficiently accelerate drawdowns to avoid default by LPs on capital commitments. LPs with higher tolerance to illiquidity then realize higher returns. With a secondary market, return persistence decreases at the GP level but persists at the LP level.

Keywords: Private Equity, Liquidity Premium, Secondary Markets.

*We would like to thank Naveen Khanna, Giorgia Piacentino, Bogdan Stacescu, Paolo Volpin, Guillaume Vuillemey (discussants), Ulf Axelson, Bo Becker, Marcus Opp, Morten Sørensen, and seminar participants at Copenhagen Business School, INSEAD, the Swedish House of Finance, TUM, EIEF, the BGSE Summer Forum, the 6$^{\text{th}}$ Cambridge Corporate Finance Theory Symposium, the 15$^{\text{th}}$ CSEF-IGIER Symposium on Economics and Institutions, the 2019 LBS Private Equity Symposium, the EFA Meeting in Lisbon, the FTG Meetings in Madrid and Pittsburgh, and the FIRS Conference in Savannah for helpful comments. All remaining errors are our own. Per Strömberg would like to thank the NASDAQ Nordic Foundation and the Söderberg Professorship in Economics for financial support. David Robinson would like to thank the Bertil Danielsson Professorship for financial support.

†Stockholm School of Economics. E-mail: vincent.maurin@hhs.se.

‡Duke University, NBER. E-mail: davidr@duke.edu.

§Stockholm School of Economics, CEPR and ECGI. E-mail: per.stromberg@hhs.se.

1 Introduction

Private Equity (PE) firms are stewards of other people's capital. They commonly raise capital through fixed-life, closed-end funds organized as limited partnerships, in which the General Partners (GPs) – the employees of the private equity firm itself – receive capital pledges from outside investors, known as Limited Partners (LPs). These commitments are drawn down over time by the GP as investments are identified. Once LPs have committed capital to a private equity fund, they are exposed to two sources of illiquidity. First, LPs have to ensure they have funds available to meet GP drawdowns. Second, PE investments take time to mature and LPs typically wait for several years to realize a positive return on their capital (inducing a cash-flow pattern known as the J-curve). In addition, net cash flows to LPs are pro-cyclical, making PE particularly illiquid during economic downturns (Robinson and Sensoy 2016). While a secondary market for partnership claims has developed in recent years, this only partially alleviates LP liquidity problems, since claims trade at substantial discounts during downturns (Nadauld et al. 2018). Despite its illiquidity, the PE asset class has grown significantly over the last 40 years (Döskeland and Strömberg 2018).

The growth of the sector has spurred a large empirical literature identifying a number of puzzling empirical regularities. There is little variation in partnership terms (Robinson and Sensoy 2013) despite large differences in performance both across managers and over time (Kaplan and Strömberg 2009, Da Rin et al. 2012). While the asset class has outperformed public markets on average over the last thirty years, the excess performance is countercyclical and particularly high when PE fundraising is low.[1] In addition, there is evidence of return persistence at the GP level, with the performance of a past fund predicting the performance of future funds raised by the same GP.[2] Persistence has also been documented at the LP level, where certain institutional investors have consistently generated higher returns in their PE portfolios than others.[3]

[1]For recent evidence, see Harris et al. 2014a and Robinson and Sensoy 2016.

[2]See Kaplan and Schoar (2005) and Harris et al. (2014a).

[3]See Lerner et al. (2007), Dyck and Pomorski (2016) and Cagnavaro et al. (2018). Whether persistence holds across LP types (pension funds, endowments, etc.) or at the individual level is open to debate (see Sensoy et al. 2014 and Cagnavaro et al. 2018).

This paper develops a parsimonious model of private equity to explain these empirical regularities. Our analysis builds on two essential features of the market. First, due to moral hazard considerations, GPs must be given incentives to properly manage investments, as in Holmström and Tirole (1997). Second, private equity investments are illiquid, and LPs differ in their tolerance to illiquidity. Moral hazard and investment illiquidity can rationalize not only the fund structure, but also the rigidity in contract terms, the excess returns to the asset class, as well as the return persistence at the GP and LP level – all without resorting to irrationality, asymmetric information or heterogeneous skill.

The commonly observed private equity fund structure emerges as a contractual solution to the agency conflict between LPs and GPs. It is optimal for LPs to commit capital for a series of investments rather than on a deal-by-deal investment basis: this makes it easier to provide incentives for GPs to add value to their investments. The compensation to GPs is a function of the overall performance of the fund and resembles the carried interest given to fund managers.[4] This optimal investment structure nevertheless leaves an unresolved commitment problem for LPs who are exposed to an aggregate liquidity shock. When hit by this shock, they may wish to default on their capital commitment to preserve liquidity for other purposes. From a GP's point of view, this friction generates an extra illiquidity cost beyond the compensation required by LPs to hold illiquid assets.[5]

In our model, investment is scalable and GPs face a trade-off between fund size and carried interest share. When PE investment becomes more attractive to LPs – for instance because underlying PE investment opportunities improve – we show that GPs respond by increasing fund size rather than their fees, consistent with the observed stickiness in carried interest.[6] The amount of external LP capital that GPs can raise is constrained by the co-investment GPs are able to make in their own funds. Following a successful fund, GP will have accumulated more wealth thanks to the performance fees they have received. They can

[4]This result is a variant of the benefits of cross-pledging with delegated monitoring, as in Diamond (1984).

[5]As we explain below, these liquidity costs are crucial to rationalize some empirical observations in PE but in the interest of clarity, we first review the results that do not depend on these features.

[6]Our model also suggests an explanation for the fact that there is more dispersion in carried interest among VC funds compared to buyout funds, since the VC investment technology is likely to be less scalable (Metrick and Yasuda 2010, Robinson and Sensoy 2013).

then provide a larger co-investment and raise more external capital in their next fund. This can explain why LPs increase their new fund commitments in response to strong previous performance, as shown by Kaplan and Schoar (2005). A similar dynamic applies at the macro level: Following a successful vintage, GPs collectively accumulate more wealth and aggregate PE fundraising should increase.

While these findings would hold even if private equity investments were fully liquid, illiquidity plays a significant role for fundraising and expected returns. The liquidity risk faced by LPs affects GPs profit through two distinct channels. First, when LPs face a lower likelihood of a liquidity shock, and/or such liquidity shocks are less costly, the premium they require for long-term investments goes down. This lower cost of capital for GPs allows them to raise larger funds. As a result, PE fundraising as well as average fund sizes are negatively related to subsequent returns, in line with the empirical evidence in Kaplan and Strömberg (2009).[7] Second, to avoid the risk of default by LPs who may experience liquidity shocks, GPs choose to call more capital in the early life of the fund.[8] Large first investments act as collateral, ensuring that LPs stand by their capital commitments for subsequent investments. This distortion, however, reduces the incentive benefits from diversification across investments and constrains GPs to raise smaller funds at the expense of total profit.

We then analyze the effect of investor heterogeneity by introducing two types of LPs, good and bad. Good LPs face milder liquidity shocks, which decreases the premium required to invest in PE and mitigates their commitment problem. Raising capital from good LPs therefore allows GPs to run larger and more efficient funds. This creates a distinction between premium capital supplied by the good LPs and the total supply of capital available in the market. When premium capital is abundant, only good LPs invest in private equity and expected returns are low. As the demand for LP capital grows, for instance when underlying investment opportunities improve, premium capital eventually becomes scarce. GPs then start raising capital also from bad LPs, who are more exposed to liquidity shocks.

[7]Our explanation complements the evidence reported by Kaplan and Stein (1993) and Gompers and Lerner (2000) in favor of a competition effect whereby low returns are driven by high deal valuations in PE boom periods. Brown et al. (2018) illustrates the difficulty that LPs face avoiding this cyclicality.

[8]Ljungqvist et al. (2017) show that GPs sometimes accelerate their draw-downs of LP commitments as a function of both GP characteristics and market conditions.

A key finding is that good LPs earn higher returns in equilibrium than bad LPs when they are both present in the market. The ability of good LPs to withstand liquidity shocks allows GPs to run more profitable funds by avoiding inefficient acceleration in drawdowns. Some GPs will therefore choose to cater only to good LPs by offering a higher expected return in their fund, while restricting access to bad LPs (these higher-returning PE funds will become "oversubscribed"). In equilibrium GPs will be indifferent between offering a high expected return to good LPs and a lower expected return to bad LPs. We therefore provide an explanation for why some GPs systematically generate higher returns to their LPs, and why some LPs are able to invest with these GPs, solely as a function of their different tolerance to liquidity shocks. In an extension, we allow GPs to have different skill levels and show that return persistence at the fund level also arises. Interestingly, positive assortative matching between skilled GPs and good LPs is not always the equilibrium outcome.

The final step of our analysis introduces a secondary market for LP investments. When an aggregate liquidity shock hits the economy, bad LPs can sell their partnership claims to good LPs with higher tolerance for illiquidity. This exit option allows investors to fully or partially realize the value of their investment before maturity. This *liquidity effect* lowers the return required by bad LPs to commit capital to PE funds. In our model, the secondary market price for PE claims is endogenous and depends on the amount of liquidity available to good LPs. When the equilibrium amount of liquidity is low, secondary claims trade at a discount. This discount arises endogenously to compensate buyers for providing liquidity in the secondary market, consistent with the empirical findings by Albuquerque et al. (2018) and Nadauld et al. (2018). Good LPs now have an incentive to save cash rather than invest in the primary market, in order to profit from expected future secondary market discounts. In contrast with the liquidity effect for bad LPs, this *opportunity cost effect* increases the expected return required by good LPs to commit capital in the primary market.

We first show that the *liquidity effect* increases the size of the primary market. By providing liquidity in the secondary market, good LPs facilitate entry of bad LPs in the primary market since the cost of illiquidity goes down for these investors. In addition, the secondary market reduces the default risk of these investors who can exit through a sale rather than through a default. As a consequence, bad LPs drive down required returns in the primary market, which increases GP profits. On the other hand, the *opportunity cost effect* leads good LPs to allocate more of their capital to the secondary market. We show that these two effects combined can generate a radical change in the investor base of private equity funds. In equilibrium, good LPs may entirely focus on the secondary market and GPs only raise capital in the primary market from bad LPs.

A direct implication of our results is that segmentation between different funds in the primary market may disappear with the introduction of a secondary market. GPs now typically only raise one type of fund instead of two types of funds with different returns. Hence, our model suggests that the decline in GP performance persistence for buyout funds, documented by Harris et al. (2014a), may be explained by the growth of the secondary market. While expected return differences across GPs should decrease, our model implies that the performance persistence among LPs should remain even when segmentation disappears in the primary market. By focusing on the secondary market, good LPs earn higher returns when claims trade at a discount during illiquid periods. In addition, even if good and bad LPs would invest in the same funds in the primary market, good LPs would still realize higher returns, since they never have to sell in the secondary market at a discount.

Although our model exhibits neither investor irrationality, nor asymmetric information, nor learning about GP skill, these features may well be important in practice. Our aim is rather to provide a benchmark model with rational and fully informed market participants against which documented empirical findings in the PE market can be assessed. In this sense, we hope to provide a private equity counterpart to the Berk and Green (2004) model of public equity mutual funds, highlighting the role of liquidity. The stylized structure of our model should also make it applicable to delegated portfolio management in other illiquid asset classes, such as infrastructure, private credit, and real estate funds.

The remainder of the paper is structured as follows. Section 2 reviews related work on private equity. We lay out the basic model in Section 3. Section 4 analyzes the optimal fund structure and Section 5 adds heterogeneity in LP types. We introduce the secondary market in Section 6. Section 7 concludes.

2 Literature Review

Starting from Jensen (1989) and Sahlman (1990), the economic structure of private equity partnerships has been interpreted as the solution to agency problems arising from delegated asset management. Our explanation for why GPs invest through funds is very similar to Axelson et al. (2009): investing through funds rather than deal-by-deal creates some "inside equity" which makes it less costly to incentivize the GP. Their analysis focuses on the role for third-party debt financing alongside PE capital to mitigate over-investment when project quality is private information. In contrast to Axelson et al. (2009), we consider a moral-hazard problem for GPs and we endogenize fund size. More importantly, we study the consequences of illiquidity on expected returns and the role of the secondary market.

Several papers provide models of the excess return of private equity over public equity and its implications for portfolio choice. Sørensen et al. (2014) and Giommetti and Sørensen (2019) investigate the illiquidity cost of private equity to investors in dynamic portfolio-choice models. In their paper, the cost of private equity is that it exposes a risk-averse LP to additional uninsurable risk. Phalippou and Westerfield (2014) also solve a dynamic optimal portfolio allocation problem for a risk-averse investor. Similar to Sørensen et al. (2014), the cost of illiquid assets arise from suboptimal diversification, but they add the feature that fund capital calls are stochastic. They also consider the possibility of LP defaults and secondary market sales, although they take the discount in the secondary market to be an exogenous parameter. In contrast to these papers, we provide an equilibrium model of delegated portfolio management, where we endogenize PE fund and compensation structures as well as equilibrium returns in the primary and secondary markets.

Similar to us, Haddad et al. (2017) explain variation in buyout activity as a result of time-varying risk premia in an agency framework. As in the papers mentioned above, the excess return on private equity compensates risk-averse investors for holding an undiversified portfolio, which in turn is necessary to provide incentives for adding value to the investment.[9] Haddad et al. (2017) argue that this compensation increases as the overall market risk pre-

[9] Ewens et al. (2013) use a similar mechanism to rationalize the high observed required rates of return that GPs use for evaluating PE investments.

mium increases, leading to pro-cyclical fundraising activity. Their model does not distinguish between LPs and GPs, but focuses on the relationship between PE investors and their portfolio companies. We analyze the frictions between GPs and LPs and we model the liquidity premium as a compensation to LPs. We also provide an explanation for return persistence.

Hochberg et al. (2014) provide a theory to explain the documented return persistence for private equity funds. In their model, LPs learn the skill of the GPs in which they invest over time, leading to informational holdup when GPs raise their next fund.[10] In contrast, our model can rationalize GP performance persistence without asymmetric information or differences in skill, as a rent provided to the most liquid LPs for providing capital. Our model incorporates a secondary market and this feature can explain why LP return persistence remains, despite the fact that GP return persistence seems to have weakened over time.

Similar to us, Lerner and Schoar (2004) argue that GPs have preference for investors with low costs of illiquidity. In their model, however, investors are uncertain about GPs' skill, and sales of LP claims in the secondary market are interpreted as negative signals of GP ability. While they do not derive an optimal fund structure, they argue that GPs endogenously limit trading of Limited Partnership claims to screen for "good" LPs. Our model allows for different funds to be raised in equilibrium offering different returns to their investors. We allow for an active secondary market and derive endogenous discounts to NAV that are not due to information asymmetries.

Finally, few papers have also modeled the secondary market for private equity fund shares. In Bollen and Sensoy (2016), a risk-averse LP allocates funds between public equity, private equity, and risk-free bonds, and has to sell PE assets at a discount if hit by an exogenous liquidity shock. They do not aim to determine primary and secondary market returns in equilibrium, but instead calibrate their model to data to determine whether observed returns and discounts can be rationalized. In our model, secondary market discounts are instead endogenously determined as a result of "cash-in-the-market pricing" when liquidity is scarce. In our model both the supply of liquidity and the long-term asset supply (private equity fund

[10] Relatedly, Glode and Green (2011) model the persistence in the returns to hedge fund strategies as a result of learning spillovers.

claims) is endogenous unlike in the model of Allen and Gale (2005) we build on. Finally, the economic mechanism leading to market segmentation in our paper, where more liquid investors focus on the secondary market in the hope of capturing "fire-sale discounts" is reminiscent of the mechanism in Diamond and Rajan (2011).

3 Model Setup

The model has three periods, denoted $t = 0, 1, 2$. The economy is populated with investors called LPs (Limited Partners) who have large financial resources and managers called GPs (General Partners) who have access to long-term investment opportunities but limited financial resources. GPs seek financing from LPs to leverage their investment skills, but need to be incentivized to exert effort in applying these skills. LPs face liquidity shocks that reduce their willingness to commit capital for long-term investments.

3.1 Limited Partners

There is a mass M of risk-neutral LPs who consume in period 1 and 2. Each LP is endowed initially with 1 unit of cash, which is storable at a net return of 0. In period 1, an aggregate liquidity shock hits the economy with probability $\lambda \in (0, 1)$. When the shock hits, LPs strictly prefer to consume early in period 1. Their preferences are given by

$$u(c_1, c_2) = \begin{cases} c_1 + c_2, & \text{with prob.} \quad 1 - \lambda \\ c_1 + \delta c_2 & \text{with prob.} \quad \lambda \end{cases} \tag{1}$$

where c_t denotes consumption in period t and $\delta < 1$ is the LPs discount factor when a liquidity shock hits. For now, we assume all LPs are ex-ante identical and have the same value of δ. We introduce LP heterogeneity in Section 5.

The liquidity shock is meant to capture an event that decreases investors' appetite for investing in long-term, illiquid assets, such as a a "flight-to-liquidity" episode[11] or a regula-

[11] See Longstaff (2004).

tory change increasing the cost for institutional investors of holding illiquid assets.[12]

These preferences imply that LPs require a return premium $r(\delta, \lambda)$ over (liquid) cash to invest in (illiquid) long-term assets. LPs are indifferent between investing in long-term assets versus holding cash if their expected payoffs are the same, that is if

$$\left[1 - \lambda + \lambda\delta\right](1 + r(\lambda, \delta)) = 1, \tag{2}$$

which implies that

$$r(\lambda, \delta) := \frac{1}{1 - \lambda(1 - \delta)} - 1 \tag{3}$$

The illiquidity premium $r(\lambda, \delta)$ increases with the probability of a liquidity shock λ and with the severity of this shock $1 - \delta$.

3.2 General Partners and Investments

There is a unit mass of risk-neutral GPs who do not discount future cash flows. GPs have an initial endowment of A units of cash in period 0. They have investment opportunities in period 0 and in period 1. Both these investments mature in period 2. Each investment returns R per unit invested in case of success and 0 in case of failure. Period 0 and period 1 investments have independent returns.

Following Holmström and Tirole (1997) we assume that GPs must exert unobservable effort for an investment to be profitable. An investment succeeds with probability p if the GP exerts effort. If the GP shirks, the probability of success of the investment is q and the GP enjoys a private benefit B per unit of funds invested. The following assumption ensures that an investment has positive NPV only when the GP exerts effort.

Assumption 1 (Shirking destroys value)

$$pR \geq 1 \geq qR + B$$

[12] For instance, after the Solvency II regulation was introduced in Europe in 2012, insurance companies were required to hold another EUR40 of balance sheet equity for every EUR100 invested in PE assets (Fitzpatrick 2011).

The leftmost term in the inequality is the investment expected payoff when the GP exerts effort. The rightmost term is the monetary payoff plus the non-monetary payoff when the GP shirks. The moral hazard problem vis à vis external investors implies that GPs must be incentivized with a claim to the investment cash flows. Following Holmström and Tirole (1997), we assume this claim needs to be large enough so as to constrain financing for GPs.

Assumption 2 (Limited pledgeability)

$$p\left(R - \frac{pB}{p^2 - q^2}\right) < 1$$

We will show that the left hand side is the maximum payoff the GPs can promise per unit of investment in a fund.[13] Assumption 2 means that this pledgeable income does not cover the investment cost. Hence, GPs cannot raise external finance without co-investing their own wealth. Finally, we assume the resources of LPs are large compared to GP wealth.

Assumption 3 (Abundant capital)

$$M \geq \frac{A}{1 - p\left(R - \frac{pB}{p^2 - q^2}\right)} - A$$

The left-hand side is the total resources in the hands of LPs. As we will show, the right hand side is the maximum amount of external financing that GPs can credibly raise. This implies that the human capital of GPs is scarce relative to the financial capital of LPs.

3.3 Partnership Contracts

In period 0, GPs compete for LPs' capital by offering investment partnership contracts. We assume that the contract terms cannot be made contingent on the realization of the liquidity shock.[14] A contract thus specifies the total fund size I (including the co-investment A by

[13]The usual version of this condition is $p\left(R - \frac{B}{p-q}\right) < 1$. Our condition is more restrictive since, as we will show, pledgeable income is higher when GPs can finance two investments jointly in a fund.

[14]While this aggregate shock is observable, contracts contingent on the realization of this shock may not be enforceable by a court. It is possible to show that our results still hold under the milder Assumption 4 to be stated in this section.

the GP), the share $x \in [0,1]$ of the fund resources I called by the GP for the period 0 investment, and the compensation schedule of the GP

$$\Big\{ w(y) \mid y \in \{0, R(1-x), Rx, R\} \Big\}, \qquad 0 \leq w(y) \leq y \tag{4}$$

The compensation schedule specifies the fee to the GP per unit of investment for each of the four possible cash flows of the fund. For instance, cash flow Rx corresponds to a success of the first investment (share x) and a failure of the second investment (share $1-x$). When the unit cash flow is y, the total compensation of the GP is equal to the fee multiplied by the fund size, that is $Iw(y)$. In period 2, the fund cash flows are realized and distributed according to the compensation schedule.

A key friction in our model is the commitment problem of investors. LPs only provide a share x of their total capital commitment in period 0. LPs that are hit by a liquidity shock in period 1 may prefer to default on their remaining commitment share $1-x$. The cost of defaulting is that LPs lose their claim to the partnership's future cash flows.[15] To formalize this trade-off, define r_{PE} as the net expected return per unit of fund capital invested. We will show that r_{PE} can be derived directly from the partnership contract features. If the liquidity shock hits in period 1, LPs will choose to honor their commitment if

$$\delta(1 + r_{PE}) \geq 1 - x \tag{5}$$

The right hand side of (5) is the benefit to LPs from avoiding the second capital call $1-x$. The left hand side is the cost of defaulting, equal to the expected value of a unit claim $1 + r_{PE}$ discounted at rate $1/\delta - 1$. LPs make the second capital call if the cost of defaulting exceeds the benefit.

The no-default constraint (5) implies that GPs may have to invest a large share x of the fund capital in period 0 to make sure LPs do not default in period 1. Alternatively, GPs

[15] Quoting from Banat-Estañol et al. (2017), "*default penalties [for LPs] are often written as long lists of punishments, ranging from relatively mild to very severe, implying the loss of some or all of the profits and the forfeiture of the defaulter's entire stake in the fund*". See also Litvak (2004).

could design the partnership such that equation (5) is violated. When a liquidity shock hits, LPs would not honor their commitment but the GP would become the sole claimant of the first investment. To simplify the analysis, we rule out this possibility.

Assumption 4 (No downsizing)
A GP with contract $(I, x, \boldsymbol{w})$ *must invest* $I(1-x)$ *in period* 1.

This assumption is meant to capture costs associated to downsizing the fund in period 1. For example, the GPs may face fixed costs (wages, searching for deals) proportional to the size of the fund. In a dynamic extension of this model, the GP could suffer reputation costs if he is unable to carry out a planned investment. We further discuss the role of this assumption after Proposition 1. Figure 1 summarizes the timeline of the model.

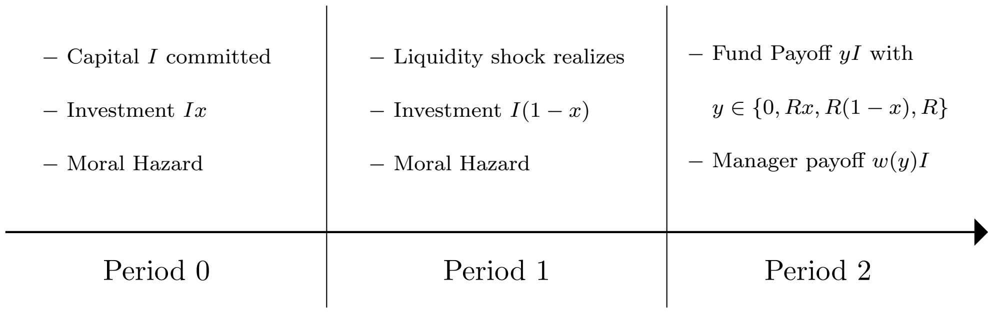

Figure 1: Timeline

4 Equilibrium PE Fund

4.1 Optimal Fund Design

Our analysis in this section delivers three main results that capture institutional features of private equity. First, GPs face a trade-off between fee and size. In our model with scalable investments, it is optimal to maximize fund size while keeping fees just high enough to incentivize the GP. Second, GPs can raise larger funds and increase profit when LPs are less sensitive to liquidity risk. Finally, our model rationalizes the fund structure whereby LPs commit capital for a series of investments rather than on an investment-by-investment basis.

4.1.1 The Trade-off Between Fees and Fund Size

Let us consider a partnership between a GP and LPs. Take as given the share of capital $x \in [0,1]$ allocated to the first investment, chosen such that the no-default constraint (5) holds. Under Assumption 1, the GP should exert effort on both investments. This requirement defines a set of incentive compatibility constraints, and we denote by $\underline{W}_x$ the minimum expected compensation to the GP (per unit of fund capital) that ensures these constraints are met. Since the GP should exert effort in each period, both the first-period investment xI and the second-period investment $(1-x)I$ have an expected return of pR per unit invested, and the expected return of the fund is equal to pRI. For an expected fee $W \geq \underline{W}_x$, the total expected payoff net of fees to LPs when investing in a fund of size I is given by

$$(pR - W)I \tag{6}$$

We can now relate the fund size I and the expected fee W to the expected return r_{PE} that LPs earn in a fund. The contribution of LPs to a fund is equal to $I - A$ and GPs contribute the remaining capital A. Since LPs' total payoff is equal to $(pR-W)I$, LPs expected return r_{PE} on their investment is given by

$$1 + r_{PE} := \frac{(pR - W)I}{I - A} \tag{7}$$

Intuitively, LPs will invest in a fund if the expected return r_{PE} weakly exceeds their break-even rate for long-term investment $r(\lambda, \delta)$. Given Assumption (3) that LPs' capital is abundant, r_{PE} will be equal to $r(\lambda, \delta)$ in equilibrium, as we show below.

GPs face a trade-off between the expected fee W and fund size I to deliver a given return on capital r_{PE} to their LPs. Suppose that a GP considers raising more external capital $I - A$. He must then reduce the expected fee W for equation (7) to hold. Similarly, if a GP raises the fee W, the amount of external capital raised, $I - A$, will have to be reduced. To derive the solution to this trade-off, let us write the total profit earned by a GP as:

$$\Pi_{GP} = \max_{W \in [\underline{W}_x, pR]} WI \qquad \text{subject to} \quad (7) \tag{8}$$

where the lower bound $\underline{W}_x$ was defined as the minimum expected fee that GPs can charge if they exert effort. Note that $W = pR$ implies that the GP receives all payoffs from the fund, but equation (7) shows that this is only possible if there is no external fund capital raised ($I = A$). In our model, where investment is perfectly scalable, this trade-off has a simple solution: GPs charge the minimum expected fee to maximize fund size.

Lemma 1 (Fee Size Trade-off)
GPs raise external capital if and only if the LPs' break-even rate is low enough, that is if

$$r(\lambda, \delta) < pR - 1 \tag{9}$$

If condition (9) holds, the equilibrium return on a private equity commitment is

$$r^*_{PE} = r(\lambda, \delta). \tag{10}$$

GPs then charge the minimum feasible expected fee $W^* = \underline{W}_x$.

Since capital supply is abundant by Assumption 3, the equilibrium return on a private equity commitment r^*_{PE} is equal to the investors' break-even rate for long-term investments $r(\lambda, \delta)$. Hence LPs are indifferent between holding liquid assets and committing capital to private equity funds in equilibrium. When condition (9) does not hold, LPs would not invest in private equity even if they could manage these investments themselves. In this case, the cost of external finance faced by GPs is too high and they only invest their own wealth A.

To see why GPs choose fund size over fees, we can use equation (8) and (10) to rewrite the expected profit of GPs as:

$$\Pi_{GP} = pRI - (1 + r(\lambda, \delta))(I - A)$$

This payoff is equal to to the total fund cash flows net of the total cost of external financing. When condition (9) holds, it is optimal to maximize the fund size I and GPs reduce the expected fee they charge to its minimum feasible value $\underline{W}_x$.

To further characterize the equilibrium, we must first derive the incentive compatible compensation schedule that minimizes the expected fee for a given value of x. The final step is to find the investment split $(x, 1-x)$ that minimizes this fee $\underline{W}_x$ under the no-default constraint (5).

4.1.2 Minimum Fee

We first show that there is an intermediate range of values of x that allows the GP to minimize the expected fee they charge. This benefit from diversifying the fund capital justifies why GPs and LPs contract for a series of investments rather than on a deal-by-deal basis. Second, we show that the commitment problem of LPs may limit these benefits.

We first derive the minimum expected fee $\underline{W}_x$ charged by GPs, focusing on the case $x \in [1/2, 1]$ without loss of generality. Under risk-neutrality, it is a well known result that the GP should be paid only after the outcome that is the most informative about effort exertion. Since a success of two independent investments is always (weakly) more informative about effort exertion that a success of a single investment, we have

$$w(0) = w(Rx) = w(R(1-x)) = 0 \tag{11}$$

that is the GP should be paid only if both investments succeed.[16] The incentive constraint for the GP is then given by the following inequality:

$$p^2 w(R) \geq \max\left\{q^2 w(R) + B, pqw(R) + Bx, pqw(R) + B(1-x)\right\} \tag{12}$$

where the payoffs on the right hand side correspond respectively to the case where the GP never exerts effort, exerts effort only on the second investment and exerts effort only on the

[16] We state formally the informativeness condition in the proof of Proposition 1.

first investment. In each case, the GP receives private benefits proportional to the fraction of the investment for which he shirks. Note that the probability of a joint success of the investments is reduced to pq (resp. q^2) when shirking on one (resp. two) investments. By Lemma 1, the GP seeks to minimize the fee charged to LPs. This implies that the IC constraint (12) binds. Since we focus on the case $x \geq 1/2$, the optimal compensation of the GP after a joint success is given by

$$\underline{w}(R) := \begin{cases} \frac{B}{p^2-q^2} & \text{if} \quad x \in \left[\frac{1}{2}, \frac{p}{p+q}\right] \\ \frac{B}{p(p-q)}x & \text{if} \quad x \in \left[\frac{p}{p+q}, 1\right] \end{cases} \tag{13}$$

and the minimum expected fee is simply given by $\underline{W}_x = p^2\underline{w}(R)$. This expected fee is minimal when $x \in \left[\frac{1}{2}, \frac{p}{p+q}\right]$ since then, the diversification benefits are maximal. Figure 2 illustrates this result by plotting $\underline{W}_x$ as a function of x where the region $[0, 1/2]$ is obtained by symmetry.[17]

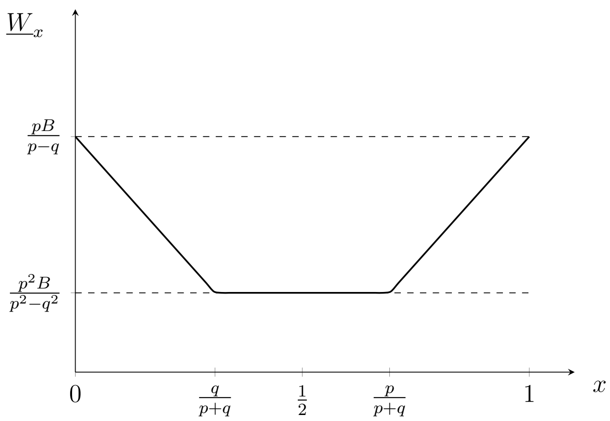

Figure 2: Minimum expected fund fee (x: *share of the first investment*)

To understand the role of the fund structure in lowering the minimum expected fee $\underline{W}_x$,

[17]The reader may observe that when R is small, it can happen that $R - \underline{w}(R) \leq Rx$. In this case, the LPs' claim would not be monotonic in the fund cash flows. This monotonicity constraint is often imposed on grounds of moral hazard to avoid misreporting of the cash flows by the manager (see for instance Innes 1990). In Internet Appendix D, we derive the optimal fund design when the monotonicity constraint on the LPs' claim is imposed. Essentially, this constraint makes it harder to give steep incentives to the GP. However, the results from Proposition 1 still hold: diversification across investments is optimal and these benefits are lower when raising funds from investors with a low value of δ.

let us consider a contract where the GP compensation is independent across investments. This is equivalent to a scheme where GPs finance investments separately. In this case, the incentive constraint needs to be met for each investment in isolation, that is

$$pw \geq qw + B \tag{14}$$

where w is the compensation per unit in case of success of an investment. The compensation schedule $\tilde{\boldsymbol{w}}$ would then be given by

$$\tilde{w}(0) = 0, \quad \tilde{w}(R(1-x)) = \underline{w}(1-x), \quad \tilde{w}(Rx) = \underline{w}x, \quad \tilde{w}(R) = \underline{w} \tag{15}$$

where $\underline{w} := \frac{B}{p-q}$ is the minimum fee the GP must charge to exert effort. The expected fee of the GP would be equal to $\tilde{W} = \frac{pB}{p-q}$ which is strictly higher than the minimum expected fee $\underline{W}_x$ unless $x \in \{0, 1\}$, as shown by Figure 2. The optimal fund structure provides better incentives to the GPs by tying the compensation for one investment to the payoff of the other investment, a result sometimes referred to as *cross-pledging*.

4.1.3 Investment distortions

Our analysis shows that GPs would choose x within the range $\left[\frac{q}{p+q}, \frac{p}{p+q}\right]$ in the absence of further constraints. But GPs must also ensure that LPs do not default on their second capital call. Using the result that $r_{PE} = r(\lambda, \delta)$ (Lemma 1), we can rewrite the no-default constraint (5) as

$$x \geq \hat{x}(\lambda, \delta) := 1 - \delta(1 + r(\lambda, \delta)) = 1 - \frac{\delta}{1 - \lambda(1-\delta)} \tag{16}$$

The no-default constraint implies that the first capital call must be large enough to ensure that LPs do not default on the second capital call. Intuitively, the first investment acts as collateral since LPs forfeit the proceeds from this investment when they default in period 1. The minimal required share of capital $\hat{x}(\lambda, \delta)$ for the first investment increases with the severity of the liquidity shock $1 - \delta$, since the LPs' loss upon default, that is the value they attribute to the first investment, decreases. In contrast, the commitment problem becomes less severe if the ex ante probability of the aggregate liquidity shock increases, since (16) shows that $\hat{x}(\lambda, \delta)$ decreases with λ. This result arises because the expected return $r(\lambda, \delta)$ that GPs have to promise LPs in period 0 increases with the likelihood of a liquidity shock, λ. This, in turn, increases the cost for LPs of giving up the first investment period 1, when the liquidity shock has already been realized.

Given the result of Lemma 1, the optimal fund structure is easy to derive. Since GPs seek to minimize the expected fee $\underline{W}_x$, they choose a value of x in the region $[\frac{q}{p+q}, \frac{p}{p+q}]$ as long as it is possible to satisfy the no-default constraint (16). Otherwise, since $\underline{W}_x$ is increasing over $[\frac{p}{p+q}, 1]$, the optimal choice of x is pinned down by the binding no-default constraint (16). The following Proposition formalizes these results.

Proposition 1 (Optimal Fund Structure)
The compensation to GPs is equal to $W^* = p^2 \underline{w}(R)$ where $\underline{w}(R)$ is given by equation (13). The first capital call is given by:

$$x^* = \begin{cases} x \in [\max\{\underline{x}, \hat{x}(\lambda, \delta)\}, \bar{x}] & \text{if} \quad \delta \geq \hat{\delta}(\lambda) \\ \hat{x}(\lambda, \delta) & \text{if} \quad \delta < \hat{\delta}(\lambda) \end{cases} \tag{17}$$

where $\underline{x} = \frac{q}{p+q}$, $\bar{x} = 1 - \bar{x}$, $\hat{x}(\lambda, \delta)$ is given by equation (16) and

$$\hat{\delta}(\lambda) := 1 - \frac{p}{p + (1-\lambda)q} \tag{18}$$

Proposition 1 shows that the capital of a private equity fund should optimally be deployed over two investments rather than one. We showed that the expected fee charged by GPs is minimized when choosing an intermediate value $x \in [\underline{x}, \bar{x}]$ for the share allocated to the first investment.[18] The benefits of diversification are not due to risk-sharing motives or complementarity between investments, but are instead a way to incentivize the GPs at a

[18]The fact that the exact value of x is not pinned down (even though it could in principle be part of the contract) is actually consistent with real-world PE partnership agreements, which specify investment

lower cost. With two independent investments, LPs receives two independent signals about effort instead of one. By tying their compensation to the joint outcome of these investments, GPs can lower their fee per unit of total investment.[19] This enables GPs to raise larger funds and increase their total profits, as shown by Lemma 1.

The key result from Proposition 1 is that the limited commitment of LPs may induce GPs to inefficiently distort the fund structure. GPs need to call and invest enough capital in period 0 to avoid default by their LPs on period 1 capital calls. When liquidity shocks are really severe, that is when $\delta \leq \hat{\delta}(\lambda)$, the minimum share to be called $\hat{x}(\lambda, \delta)$ lies outside the optimal region $[\underline{x}, \bar{x}]$. GPs are forced to call "too much" capital early and do not reap the full incentive benefits of diversification. This distortion increases the expected fee and reduces fund size, leading to a lower profit for the GP.[20]

GPs thus incur a cost when raising capital from LPs who face severe liquidity shocks, in addition to the liquidity premium they have to pay. The expression for the threshold $\hat{\delta}(\lambda)$ in equation (18) shows that this cost of commitment decreases with p/q, and with λ. When p/q is large, diversification benefits arise for a larger range of values of x, as shown by equation (13). This allows GPs to increase the first capital call x in order to avoid default by LPs while keeping the fee minimal. When the probability λ of a liquidity shock increases, the cost of capital $r(\lambda, \delta)$ and thus the return to LPs goes up. Since the foregone profits would be higher, it is more costly for LPs to walk away.

concentration limits rather than specific capital deployment schedules (Gompers and Lerner 1996). For buyout funds, the investment in any given deal usually cannot exceed 20% of the total fund commitment.

[19] This insight about the benefits of diversification in the context of delegated monitoring is originally due to Diamond (1984) and Laux (2001). The model of Axelson et al. (2009) rely on a similar argument to derive the compensation structure of GPs under asymmetric information about investment quality.

[20] Anecdotal evidence from the financial crisis suggests that GPs indeed accelerate investment in response to concerns about LP defaults. In a May 15, 2009 article for *PE Hub, Buyout Insiders*, Erin Griffith observes:

> *How are general partners avoiding that potentially messy situation* [default by investors]*? Some funds have drawn up to 20% of their capital upon closing, panelists at the Masterclass said. Its a way for GPs to make sure their investors have skin in the game from the start.*

4.2 Comparative Statics

Using Proposition 1, we can derive the equilibrium fund size and the GP's profit. Distinguishing between the two cases in Proposition 1, we have

$$W^* = \begin{cases} \frac{p^2 B}{p^2 - q^2} & \text{if} \quad \delta \geq \hat{\delta}(\lambda) \\ \left[1 - \delta(1 + r(\lambda, \delta))\right] \frac{pB}{p-q} & \text{if} \quad \delta < \hat{\delta}(\lambda) \end{cases} \tag{19}$$

$$I^* := \frac{A}{1 - \frac{1}{1+r(\lambda,\delta)} \left[pR - W^*\right]} \tag{20}$$

$$\Pi^*_{GP} := W^* I^* \tag{21}$$

The fund size I^* is linear in the GP's own contribution to the fund A. The ratio between the two variables is sometimes called the equity multiplier. This equity multiplier is large when the cost of capital $r(\lambda, \delta)$ is low or when the pledgeable income per unit of investment, equal to $pR - W^*$ is high. Both the probability of a liquidity shock λ and the severity of this shock $1 - \delta$ increases the cost of capital for GPs. When $\delta < \hat{\delta}(\lambda)$, a more severe liquidity shock further reduces the pledgeable income since the investment schedule is suboptimal. The following results formalize these observations.

Corollary 1 (Comparative Statics)
Fund size I^* and GPs' profit Π^*_{GP} are decreasing in LPs' liquidity risk λ and in the severity of the liquidity shock $1 - \delta$. They are increasing with the investments' payoff R and with the probability of success q when $p - q$ is kept constant.

The effect of an increase in the probability of the liquidity shock λ is not obvious. A higher value of λ increases the cost of capital $r(\lambda, \delta)$ but this increase in $r(\lambda, \delta)$ indirectly relaxes the no-default constraint (16). However, the first effect dominates and size and profit decrease with λ. This finding confirms the result implicit to Proposition 1 that GPs would rather distort investment than increase returns to alleviate the commitment problem.

The effect of an increase in the return R in case of success is intuitive. When investments are more profitable, the total payoff to investors net of fees per unit of investment goes up.

GPs can then scale up their funds to bring back investors return at their break-even rate. When varying the probability of success q under shirking, we fix the difference $p - q$. This allows us to isolate the contribution from a better economic environment (higher q) from that of having more efficient GPs (higher p). When p increases but q does not, investment becomes more profitable and incentives are also cheaper to provide. Both these effects contribute to an increase in size and profit.

4.3 Empirical Relevance

Our model can rationalize several empirical relationships between PE fund compensation, fundraising, and returns that have been documented in the literature. Fees in private equity vary remarkably little across funds and over time, especially when it comes to the carried interest ($\underline{w}(R)$ in our model), where 94% of the PE funds in Robinson and Sensoy (2013) have a carried interest of exactly 20%. Instead, average funds size increases in periods of high aggregate fundraising. Our model indeed suggests that GPs should increase fund size rather than fees. Kaplan and Schoar (2005) also find that PE firms raise the size of their funds when previous fund performance has been relatively strong. In our model, successful GPs will have earned higher carried interest and will therefore have more wealth A to invest in their next fund. This, in turn, increases the amount of capital $I - A$ they can raise from LPs.[21] Finally, Kaplan and Strömberg (2009) show that funds raised during strong fundraising periods have lower returns. In our model, GPs indeed raise more capital when the compensation for illiquidity required by investors $r(\lambda, \delta)$ and thus returns are low.

A straightforward extension of our model could also help explain some systematic differences between buyout and VC funds that have been documented in the literature. Our baseline result that GPs prefer to increase size rather than fees relies on the technological assumption that investment is perfectly scalable, that is R or p do not depend on I. This assumption is more plausible for buyout funds, where a manager who raises a larger fund can simply acquire larger portfolio companies using a similar investment approach. In con-

[21] Stretching the theory, one could imagine that strong performance in the past fund would lead LPs to increase their expectation of R and/or p, which would also lead to an increase in I.

trast, a VC manager investing in early-stage start-ups cannot as easily scale up the amount invested in any given company, since start-ups are almost by definition bounded in size.[22] With limited investment scalability, our model predicts that successful GPs should respond by increasing fees W when they cannot increase the fund size I. Consistent with this prediction, Robinson and Sensoy (2013) shows that the variation in carried interest is much lower in buyout funds, where only 1% of funds have carried interest above 20%, compared to VC funds where this number is 10%.

We conclude this section by highlighting the dual role of liquidity risk which is specific to our model. Investors that are less sensitive to liquidity risk have a lower break-even rate for long-term investments. This decreases the cost of capital for GPs who raise larger funds and earn more profit. However the cost of capital does not fully capture the cost of illiquidity for GPs. Among two hypothetical group of investors with different liquidity exposure (λ, δ) but with the same break-even rate $r(\lambda, \delta)$, GPs prefer LPs for which $\delta > \hat{\delta}(\lambda)$, that is LPs who can better commit to late capital calls. The default risk on capital calls for "bad" LPs translates into an investment distortion for GPs who must ultimately raise smaller, less profitable funds. We will show in the next section that this second feature explains why "good" LPs can earn higher returns in equilibrium.

4.4 Alternative fund structures

At this point, it is useful to discuss alternative ways to mitigate the LP commitment problem, and explain why they do not dominate the fund contract we propose; either because they are not feasible, or because they impose additional costs on the GPs. Appendix B provides details and proofs of the claims below.

First, it may seem that the GP can avoid the commitment problem of LPs by calling extra capital in period 0. He would then hold as cash any capital called in excess of the

[22] This reading is supported by Metrick and Yasuda (2010) who find that buyout managers build on their prior experience by increasing the size of their funds faster than VC managers do, and conclude that the buyout business is more scalable than the VC business. The limited scalability of VC is also supported in Kaplan and Schoar (2005), who find that the sensitivity of fund size to past performance is significantly stronger in buyout compared to VC.

amount needed for the first contractual investment. However, we show in Appendix B.1 that this arrangement is not incentive compatible. The GP would then deviate by investing all the capital called in period 0 and shirk on the investment. Hence, GPs cannot credibly solve the commitment problem of their LPs by drawing down more capital in period 0.[23]

Second, instead of drawing down on the ex ante LP commitment to finance the second investment, GPs could potentially raise fresh capital from new investors in period 1. As we show in Appendix B.2, however, the expected cost of capital is higher in period 1 than the ex-ante cost of capital in period 0. When this cost is high, GPs would still prefer the fund structure we propose, despite the resulting investment distortions. To simplify the analysis, we rule out period 1 financing entirely in the main text.[24]

Third, Assumption 4 rules out contracts such that LPs optimally default on their commitment after a liquidity shock. Without this assumption, GPs might prefer to offer contracts that induce default, despite this leading to a drop in total investment in bad times. The reason GPs might prefer this contract, even though the fund gets downsized, is that they become the sole claimants of the existing investment when LPs default. By imposing additional costs from downsizing, we rule out this possibility. As we explained, downsizing may be difficult in practice since GPs are likely to face substantial fund management costs (salaries, searching for deals) that are proportional to the initial size of the fund. There might also be reputational costs for a GP who is not able to invest all the committed capital.

[23]GPs would be unable to over-invest if the additional cash was kept in an escrow account only available to the GP in period 1. However, this option would be costly for LPs if their outside option were better than cash. Then, GPs would need to compensate LPs for the opportunity cost of tying up their capital in a low return escrow account. A straightforward extension of our model, with a liquid outside investment option dominating cash for LPs, would allow us to flesh out this insight.

[24]In addition, there are reasons to believe this solution would be infeasible in practice. Such a contract would need to specify an ex ante contract in which payouts to GPs and LPs are contingent on actions and investments being undertaken by third parties ex post, who are not bound to this initial agreement. Such complex contracts would likely be impractical and difficult to enforce in practice, not the least because PE funds typically make 10-20 investments during their investment period. Moreover, raising new capital would divert the GP's attention from ongoing investments. Consistent with these arguments, PE partnership contracts that *require* GPs (and make compensation contingent on) future capital raising is, to our knowledge, not observed in practice.

5 Heterogeneous Illiquidity Tolerance

In our analysis so far, GPs faced an homogeneous population of LPs with the same preferences for liquidity. In this section, we consider two classes $i \in \{L, H\}$ of LPs who differ according to the severity of the aggregate liquidity shock $1 - \delta_i$.

Assumption 5 (Heterogeneous LPs)

$$\delta_L < \hat{\delta}(\lambda) < \delta_H \tag{22}$$

According to our previous analysis, Assumption 5 implies that H-LPs are better investors than L-LPs for private equity. They require a lower rate of return, that is $r(\lambda, \delta_H) < r(\lambda, \delta_L)$ and they do not risk defaulting on capital calls. This assumption is meant to capture the heterogeneity in maturity profile, investment horizon or exposure to regulatory shocks among institutional investors in private equity. The total resources M of investors are divided between H-LPs with a share μ_H and L-LPs with a share $1 - \mu_H$. The variable μ_H thus captures the average tolerance to illiquidity in the population of investors.

Our objective it to characterize the private equity market equilibrium as a function of the share μ_H of H-LPs' capital. For simplicity, we assume that the same contract must be offered to all investors in a given fund. However, GPs can still select the type of LPs they allow in their fund.[25] Our assumption will imply that a given GP raises capital from one type of LP. We thus call i-fund a fund with i-LPs as investors.[26]

In this setting, two pairs of variables describe an equilibrium. First, we define $\alpha_i \in [0, 1]$ as the fraction of GPs who raise a i-fund. Second, we let $r_{PE,i}$ be the return on a dollar invested by a LP in a i-fund. The analysis in Section 4 showed that the expected fee W_i^* in

[25] If types were not observable, L-LPs would pretend to be H-LPs and default on their capital calls when the liquidity shock hits. However, there is substantial evidence that GPs care about the liquidity profile of the investors allowed in the fund. See for instance Lerner and Schoar (2004).

[26] If we allow GPs to raise funds with different contracts tailored to each type of LPs, funds with mixed investor composition can arise in equilibrium. Intuitively, if capital supplied by H-LPs is more expensive, GPs would try to raise the minimum amount from H-LPs that avoids the investment distortion, rather than raising capital only from H-LPs. However, our key results survive: H-LPs earn higher returns and fund segmentation emerges when μ_H is low. The formal results are available upon request.

a i-fund is given by

$$W_H^* := \frac{p^2 B}{p^2 - q^2}, \qquad W_L^* := \frac{pB}{p-q}\hat{x}(\lambda, \delta_L) > W_H^* \tag{23}$$

Given an expected return $r_{PE,i}$, the supply of capital by i-LPs is given by

$$S_i(r_{PE,i}) := \begin{cases} 0 & \text{if} \quad r_{PE,i} < r(\lambda, \delta_i) \\ S \in [0, \mu_i M] & \text{if} \quad r_{PE,i} = r(\lambda, \delta_i) \\ \mu_i M & \text{if} \quad r_{PE,i} > r(\lambda, \delta_i) \end{cases} \tag{24}$$

As we saw, risk-neutral LPs supply all their capital when the return on investing exceeds their break-even rate. Using equation (7), the demand from capital by GPs managing a i-fund can be expressed as a function of the return paid to i-LPs:

$$I_i(r_{PE,i}) - A := \frac{A}{\frac{1+r_{PE,i}}{pR - W_i^*} - 1} \tag{25}$$

The profit of a GP when managing such a fund is $\Pi_i(r_{PE,i}) = W_i^* I_i(r_{PE,i})$. An equilibrium is defined as follows:

Definition 1 (Equilibrium with heterogeneous LPs)
An equilibrium is given by returns $(r^*_{PE,L}, r^*_{PE,H})$ and a fund composition (α^*_L, α^*_H) such that:

1. *(Optimal fund choice) $\alpha_i^* > 0$ iff $\Pi_i(r^*_{PE,i}) = \arg\max \left\{ \Pi_L(r^*_{PE,L}), \Pi_H(r^*_{PE,H}) \right\}$.*

2. *(Market Clearing) For $i \in \{L, H\}$, $S_i(r^*_{PE,i}) = \alpha_i^*(I_i(r^*_{PE,i}) - A)$*

The first equilibrium requirement is that GPs only offer a given type of funds if it delivers the highest profit. It implies that L-funds and H-funds coexist in equilibrium only if they deliver the same profit.

Proposition 2 (Heterogeneous LPs)

Under Assumption 5, there exists $(\underline{\mu}_H, \bar{\mu}_H)$ *where* $0 < \underline{\mu}_H < \bar{\mu}_H < 1$ *such that*

1. *When* $\mu_H \geq \bar{\mu}_H$*, there is no* L*-fund, that is* $\alpha_L^* = 0$ *and* $r_{PE,H}^* = r(\lambda, \delta_H)$*.*

2. *When* $\mu_H \in [\underline{\mu}_H, \bar{\mu}_H]$*, there is no* L*-fund, that is* $\alpha_L^* = 0$ *and*

$$r_{PE,H}^* = \frac{\bar{\mu}_H(A + \mu_H M)}{\mu_H(A + \bar{\mu}_H M)}\left[1 + r(\lambda, \delta_H)\right] - 1 > r(\lambda.\delta_H)$$

3. *When* $\mu_H \leq \underline{\mu}_H$*, both funds exist in equilibrium.* H*-LPs earn a premium, that is*

$$r_{PE,H}^* > r_{PE,L}^* = r(\lambda, \delta_L), \qquad \textit{where} \quad W_H^* I_H(r_{PE,H}^*) = W_L^* I_L(r(\lambda, \delta_L))$$

   The fraction $\alpha_H^* < 1$ *of* H*-funds solves* $\alpha_H^*(I_H(r_{PE,H}^*) - A) = \mu_H M$*.*

The expression for the thresholds $\underline{\mu}_H$ and $\bar{\mu}_H$ can be found in the proof. Proposition 2 first shows that GPs do not raise L-funds if capital from H-LPs is abundant (Case 1). This finding is intuitive since H-LPs have a lower break-even rate $r(\lambda, \delta_H)$ than L-LPs. When μ_H is high, H-LPs collectively have enough resources to meet the demand for capital from GPs. When μ_H is intermediate, the resources of H-LPs become scarce and the market clears at an equilibrium rate $r_{PE,H}^*$ above the break-even rate of H-LPs. As long as $\mu_H > \underline{\mu}_H$, L-LPs cannot compete away these rents because the return still falls short of their own break-even rate $r(\lambda, \delta_L)$. When the share μ_H of capital available to H-LPs decreases below $\underline{\mu}_H$, the cost of capital for H-funds becomes so high that some GPs raise funds L-funds.

The second key finding is that GPs pay a higher return to their investors in a H-fund when both types of funds exist in equilibrium. This premium arises because GPs must charge a higher expected fee in a L-fund, reflecting the investment distortion compared to a H-fund. Since total profit is decreasing in the expected fee W_*, H-funds would be more profitable if $r_{PE,H}^*$ were equal to $r_{PE,L}^*$. Hence, by the optimal fund choice condition of Definition 1, an equilibrium with both types of fund can only exist if the cost of capital is higher for a H-fund, that is $r_{PE,H}^* > r_{PE,L}^*$. This premium reflects the higher willingness of GPs to pay for capital supplied by H-LPs. In H-funds, GPs can optimally diversify their investments by limiting the amount of capital called in period 0. GPs who compete for the capital provided by H-LPs pay a premium when this special capital is scarce. Figure 3 illustrates our findings about PE returns with heterogeneous investors.

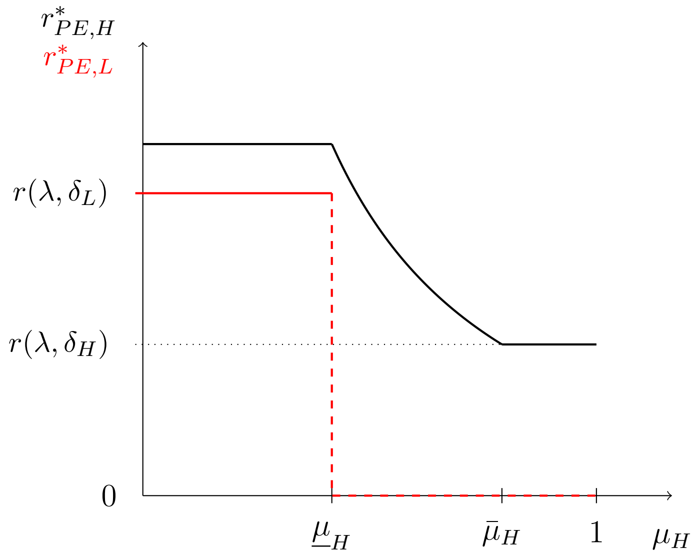

Figure 3: Expected returns from private equity investment

5.1 Limited Partner Performance Persistence

Our results resonate with the recent evidence about performance persistence for LPs. Cagnavaro et al. (2018) and Dyck and Pomorski (2016) show that some LPs or types of LPs earn consistently higher returns on their private equity portfolio. Our theory is that GPs "cherry-pick" their LPs based on their liquidity profile. The best LPs earn a higher return compensating their ability to withstand liquidity shocks. GPs value this commitment ability of LPs which allows them to run more profitable funds. Our argument contrasts with the view that some LPs may be better at selecting their GPs. In particular, only our explanation based on LP screening can explain why some GPs ration access to their funds.

5.2 Return Persistence for General Partners

Should return only persist at the investor level or also at the GP level as Kaplan and Schoar (2005) suggests? A simple extension of the model accommodates heterogeneity in GP skills as in Berk and Green (2004). Suppose that some GPs have special skills and their investment pays off R_g in case of success while for other GPs, this payoff is only $R_b < R_g$. Our previous analysis suggests that when μ_H is low, GPs compete for scarce premium capital supplied by H-LPs. GPs win the competition if they are willing to pay a higher rate for premium capital. Let us denote $\bar{r}_{PE,H}(R)$ the highest rate a GP is willing to pay for H-LPs' capital as a function of R. This rate is pinned down by the indifference condition of a GP between a H-fund and a L-fund, stated in Definition 1. We obtain

$$\bar{r}_{PE,H}(R) = \frac{W_L^*\big(pR - W_H^*\big)}{\frac{W_H^*}{1+r(\lambda,\delta_L)}\big(pR - W_L^*\big) + W_L^* - W_H^*} - 1 \tag{26}$$

where we set $r^*_{PE,L} = r(\lambda, \delta_L)$ since L-LPs always earn their break-even rate. Since $\bar{r}_{PE,H}(R)$ is generically strictly monotonic in R as we show below, it follows that one type of GP is willing to pay a higher return to raise a H-fund. Hence if heterogeneity in skills are persistent, return persistence also arises at the fund level. Perhaps surprisingly, we also show that H-LPs do not always match with the best GPs.

Corollary 2

The return $\bar{r}_{PE,H}(R)$ is increasing in R if and only if $W_H^* < 1+r(\lambda,\delta_L)$. When this condition holds, H-funds are raised by the GPs with higher absolute performance $R_g > R_b$.

Corollary 2 shows that good GPs are not always willing to pay more for premium capital than bad GPs. The intuition is that good GPs also make more profit than bad GPs when raising funds from L-LPs. Hence, they may prefer raising cheap L-funds when the marginal benefit from external financing in these funds, equal to $pR - (1 + r(\lambda, \delta_L))$ is high. When instead the pledgeable income $pR - W_H^*$ in a H-fund is high, good GPs will raise capital from H-LPs. Hence, if the condition in Corollary 2 holds, the matching result between good LPs and good GPs naturally follows. Then, when μ_H is low, H-LPs only invest in funds run by good GPs because they have a higher willingness to pay for premium capital.

This simple extension thus explains why the same funds may consistently deliver higher returns to their investors. However, our result suggests that positive assortative matching between good GPs and good LPs needs not be the equilibrium outcome. We also stress that return persistence at the fund level would disappear in our model if the only source of heterogeneity is GP skills. Better GPs would simply raise larger funds. Hence, according to our model, differences in LPs liquidity profile is a fundamental source of return persistence while heterogeneity in GP skills is not. In the remainder of the text, we return to our benchmark where all GPs have the small investment skills.

5.3 Return Premium: Comparative Statics

We provide comparative statics for the premium earned by H-LPs. Corollary 2 already shows that the effect of investment profitability is ambiguous. We study below the effect of the parameters determining the liquidity costs.

Corollary 3 (Return Premium)
When premium capital is scarce, that is $\mu_H < \underline{\mu}_H$, the premium $\frac{r^*_{PE,H} - r^*_{PE,L}}{1+r^*_{PE,L}}$:
– decreases in the probability of the liquidity shock λ.
– may increase or decrease in the severity of the shock $1-\delta_L$ for L-LPs.

When the probability of a liquidity shock or the severity of the shock increases, the break-even rate $r(\lambda, \delta_L)$ of L-LPs goes up. However, since there is not a full pass-trough to the rate $r^*_{PE,H}$ earned by H-LPs, the premium tends to go down. The overall effect of the severity of the shock $1-\delta_L$ is ambiguous because there is another countervailing force in this case. When δ_L decreases, funds raised from L-LPs become less efficient because the commitment problem of investors worsens. This increases GPs' willingness to pay for premium capital and this effect counteracts that of the imperfect pass-trough explained above.

We conclude this section noting that the commitment friction studied in Section 4 is key to explain return persistence in our model. Suppose that we modify Assumption 5 so that $\hat{\delta}(\lambda) < \delta_L < \delta_H$. This implies that the commitment problem of L-LPs is also moot. Obviously, L-LPs still have a higher break-even rate for investment than H-LPs. Our previous analysis thus implies that for high μ_H, H-LPs are the only investors in private equity. However, because the commitment problem of L-LPs has no bite, GPs have the same willingness to pay for capital from H-LPs or L-LPs. This implies that H-funds and L-funds would deliver the same return to their investors in equilibrium when μ_H is low.

6 A Secondary Market for PE Commitments

We now consider a secondary market for LPs claims. The secondary market opens in period 1 after liquidity shocks are realized. The market in period 0 is now called the primary market. When a liquidity shock hits, the secondary market allows L-LPs who invested in the primary market to sell their claim to H-LPs. These gains from trade arise because H-LPs value period 2 cash flows more under Assumption 5. A secondary claim entitles his new owner to the cash flow rights attached to \$1 of capital committed. This normalization implies that the initial LP makes the second capital call before selling the claim.[27]

The objective of this section is to determine the price of secondary market claims and the effect of the secondary market on the primary market for private equity. In normal times, there is no secondary market since all investors have the same value for period 2 cash flows. In addition, the secondary market for claims in H-funds is inactive even in bad times because no investor in the economy values the claim more than a H-LP. Without ambiguity, we call P_L the secondary market price of a claim in a L-fund when the liquidity shock hits.

We now explain the role of the secondary market on investors' choices in period 1 and period 0. First we consider trades in the secondary market in period 1. Then we consider how these anticipated trades affect investment and capital allocation decisions in period 0.

[27] We could alternative assume that the new LP makes the capital call. The price of the claim would decrease by $1 - x$, reflecting the liability acquired by the buyer. The economics are unaffected.

6.1 Secondary Market Trades

As we observed, only L-fund claims trade in the secondary market. Given their primary market commitment $S_L(r_{PE,L}, P_L)$, the supply of claims from L-LPs is given by:

$$S^{sec}(r_{PE}, P_L) = \begin{cases} S \in [0, S_L(r_{PE,L}, P_L)] & \text{if} \quad P_L = \delta_L(1 + r_{PE,L}) \\ S_L(r_{PE,L}, P_L) & \text{if} \quad P_L > \delta_L(1 + r_{PE,L}) \end{cases} \tag{27}$$

By linearity of preferences, L-LPs sell their entire participation if the price exceeds their reservation value $\delta_L(1 + r_{PE,L})$. On the demand side, H-LPs can buy claims using any resources net of their own capital commitments $S_H(r_{PE,H}, r_{PE,L}, P_L)$ in the primary market. The demand for claims is thus given by:

$$D^{sec}(r_{PE,H}, r_{PE,L}, P_L) = \begin{cases} \frac{\mu_H M - S_H(r_{PE,H}, r_{PE,L}, P_L)}{P_L} & \text{if} \quad P_L < \delta_H(1 + r_{PE,L}) \\ D \in \left[0, \frac{\mu_H M - S_H(r_{PE,H}, r_{PE,L}, P_L)}{P_L}\right] & \text{if} \quad P_L = \delta_H(1 + r_{PE,L}) \end{cases} \tag{28}$$

With linear preferences, H-LPs spend all their available resources to buy claims when the price P_L is lower than their reservation value $\delta_H(1 + r_{PE,L})$.

6.2 Primary Market Fundraising

6.2.1 GPs' Capital Demand

In period 0, GPs raise funds taking the cost of capital $r_{PE,i}$ for i-funds as given. In a i fund, GPs must also choose the share x_i of capital called in period 0. As we showed in Section 4, the choice of x is constrained by the commitment problem of LPs. The presence of a secondary market changes the no-default constraint of L-LPs who can now sell their claim. Condition (5) then writes:

$$\max\{\delta_L(1 + r_{PE,L}), P_L\} \geq 1 - x \tag{5b}$$

since a L-LP can either default or sell his claim at price P_L in the secondary market. Equation (5b) shows that this exit option for LPs can reduce the investment distortion for GPs since they can call more capital (lower x) in period 1 when P_L is high. Formally, the choice of x_i by GPs is given by

$$x_H^* = \bar{x} \tag{29}$$

$$x_L^* = \max\left\{\bar{x}, 1 - \max\{\delta_L(1 + r_{PE,L}), P_L\}\right\} \tag{30}$$

where $\bar{x}$ is the highest value of x in the optimal range for the first capital call defined in Proposition 1. By Assumption 5, the commitment problem is moot for H-LPs so x_H^* can be chosen optimally.

We now write the capital demand from GPs for each type of fund, using equation (24):

$$I_i(r_{PE,i}, x_i) - A = \frac{A}{\frac{1+r_{PE,i}}{pR - \underline{W}_{x_i}} - 1} \tag{31}$$

where $\underline{W}_x$ is the expected fee charged by GPs, given by equation (13), when calling a fraction x of the capital in period 0. The GPs' profit with a i-fund is thus given by $\Pi_i(r_{PE,i}, x_i) = \underline{W}_{x_i} I_i(r_{PE,i}, x_i)$. The secondary market price has a direct effect on the capital demand for L-funds since x_L depends on P_L by equation (30). We will show that in equilibrium the secondary market price also affects the capital demand indirectly through the return required by LPs on their capital commitments.

6.2.2 LPs Portfolio Choice in Period 0

We finally consider the portfolio choice of LPs in period 0. We saw that in the absence of a secondary market, a i-LP commits all his resources to private equity if the return offered $r_{PE,i}$ exceeds his break-even rate $r(\lambda, \delta_i)$. With a secondary market, this comparison is not relevant anymore. To describe this new trade-off, we define $u_{PE,i}$ and $u_{c,i}$ as the net return in utils from one unit of capital invested in a PE fund and stored in cash respectively. For

L-LPs, we have:

$$1 + u_{PE,L} = \lambda \max\{P_L, \delta_L(1 + r_{PE,L})\} + (1 - \lambda)(1 + r_{PE,L}) \quad (32)$$

$$1 + u_{c,L} = 1 \quad (33)$$

With a secondary market, L-LPs will sell their claim when the price P_L exceeds their reservation value $\delta_L(1 + r_{PE,L})$ for the claim. In particular, if P_L strictly exceeds this value, the expected return $r_{PE,L}$ on a capital commitment that makes L-LPs indifferent is strictly lower than the break-even rate $r(\lambda, \delta_L)$. Since L-LPs can increase their return by selling their claim in the secondary market when a liquidity shock hits, they accept a lower nominal rate to commit capital in the primary market. We call this effect the *liquidity* effect of a secondary market.

The secondary market alters the H-LPs portfolio choice in a different way. These investors do not use the secondary market as a source of liquidity for their primary market investment but rather as buyers of claims sold by L-LPs. Their returns on a PE fund investment and cash are given respectively by:

$$1 + u_{PE,H} = \lambda\delta_H(1 + r_{PE,H}) + (1 - \lambda)(1 + r_{PE,H}) \quad (34)$$

$$1 + u_{c,H} = \lambda \max\left\{\frac{\delta_H(1 + r_{PE,L})}{P_L}, 1\right\} + 1 - \lambda \quad (35)$$

Unlike for L-LPs, the return on a private equity commitment for H-LPs in (34) is not directly affected by the presence of a secondary market. These investors hold on to their claim even when a liquidity shock hits. Hence, fixing the return on the outside option to 0, the break-even rate of H-LPs would be given again by $r(\lambda, \delta_H)$. This can be seen by setting $u_{PE,H}$ to 0 in equation (34). However, the utility weighted return on cash $u_{c,H}$ may now exceed 0 since H-LPs can buy claims in the secondary market. In particular, if $P_L < \delta_H(1 + r_{PE_L})$, secondary claims are cheap from the point of view of H-LPs. Hence, the minimum return required by H-LPs would exceed their break-even rate $r(\lambda, \delta_H)$ because of the *opportunity cost* of funds to be deployed in the secondary market.

The utility weighted returns allow us to write the portfolio choice of LPs in the primary market in a simple way. LPs invest in PE funds if the return on their investment exceeds the return on cash. The capital supply in the primary market is then given by:

$$S_i = \begin{cases} 0 & \text{if} \quad u_{PE,i} < u_{c,i} \\ S \in [0, \mu_i M] & \text{if} \quad u_{PE,i} = u_{c,i} \\ \mu_i M & \text{if} \quad u_{PE,i} > u_{c,i} \end{cases} \tag{36}$$

Equation (36) is similar to its counterpart (24) but the minimum returns required by LPs now depend on the endogenous price of claims in the secondary market.

6.3 Equilibrium with a Secondary Market

An equilibrium is defined as follows:

Definition 2 (Equilibrium with a secondary market)
An equilibrium is given by returns $(r^*_{PE,L}, r^*_{PE,H})$, a fund composition (α^*_L, α^*_H) and a secondary market price P^*_L such that the following conditions are satisfied:

1. *(Optimal fund choice)* $\alpha^*_i > 0$ *iff* $\Pi_i(r^*_{PE,i}, x^*_i) = \arg\max \left\{ \Pi_L(r^*_{PE,L}, x^*_L), \Pi_H(r^*_{PE,H}, x^*_H) \right\}$.
2. *(Primary Market Clearing) For* $i \in \{L, H\}$, $S_i = \alpha^*_i (I_i(r^*_{PE,i}, x^*_i) - A)$
3. *(Secondary Market Clearing)* $D^{sec}(r^*_{PE,L}, P^*_L) = S^{sec}(r^*_{PE,L}, P^*_L)$ *if* $\alpha_L > 0$ *and* $P^*_L = \delta_H(1 + r^*_{PE,L})$ *otherwise,*

where x^*_H *and* x^*_L *are given by equations (29) and (30).*

The definition of an equilibrium builds on Definition 1 adding the secondary market clearing condition. Note that if $\alpha^*_L = 0$, no L-fund is raised in equilibrium and there is no secondary market trading. GPs who contemplate raising L-funds must still form expectations about P^*_L. In this case, the equilibrium price is set to the highest valuation for the claim among all investors in the economy. This selection device avoids coordination problems whereby GPs do not raise L-funds because they expect low secondary market prices.

It will be useful to introduce the familiar concept of a discount to Net Asset Value in the secondary market. In our model, the endogenous equilibrium discount D^* is given by:

$$D^* = 1 - \frac{P_L^*}{1 + r_{PE,L}^*} \in [1 - \delta_H, 1 - \delta_L] \tag{37}$$

A claim trades at a discount when the secondary market price P_L^* is lower than the expected value of the claim $1 + r_{PE,L}^*$. The bounds obtain because, using equations (27) and (28), the secondary market price can only clear if P_L^* lies between $\delta_L(1 + r_{PE,L}^*)$ and $\delta_H(1 + r_{PE,L}^*)$. In our model the lowest possible discount is strictly positive. When a liquidity shock hits, even the investors with the highest valuation for the claim discount period 2 cash flows. Hence, the existence of a discount follows mechanically from the assumption that $\delta_H < 1$. But we will show that the equilibrium discount can increase above this baseline value.[28]

Before stating the main proposition, we show that in the presence of a secondary market, L-LPs require a lower return on their private equity investment than H-LPs. For a given discount D, we define this minimum return $\underline{r}_{PE,i}(D)$ as the value of $r_{PE,i}$ such that the return on cash $u_{c,i}$ is equal to the return on private equity $u_{PE,i}$ using equations (32)-(35).

Lemma 2 (Cost of Capital)
Let $D \in [1 - \delta_H, 1 - \delta_L]$ be the discount to NAV in the secondary market. The difference $\underline{r}_{PE,H}(D) - \underline{r}_{PE,L}(D)$ in the minimum rate of return required by H-LPs compared to L-LPs is equal to 0 when $D = 1 - \delta_H$ and it is increasing in D.

Remember that without a secondary market, the opposite result holds since H-LPs break-

[28] In practice, the expected return is not known and must be estimated by GPs. Empirical studies like Albuquerque et al. (2018) show that PE fund claims sometimes trade at a premium over NAV (a negative discount). Our slightly different definition does not allow for negative discounts. The usual concept of discount to NAV only applies to the drawn portion of the commitment while we normalize by the expected values of all the commitments. In our framework, the standard definition would read

$$\hat{D} = 1 - \frac{P}{x(1 + r_{PE})}$$

Proposition 3 then shows that premia to NAV arise in equilibrium under this definition. Essentially, the standard concept of discount to NAV prices the remaining capital calls at cost while we use the expected *fair* value. Our measure can also be related to that used in Nadauld et al. (2018). In our model, P is the "return to a seller" while $(1 + r_{PE})/P$ is the "the return to a buyer".

even rate $r(\lambda, \delta_H)$ lies below the break-even rate $r(\lambda, \delta_L)$ of L-LPs. To understand why this inequality is now reversed, observe that the secondary market lowers the minimum rate required by L-LPs through the *liquidity effect*. Simultaneously, the minimum rate required by H-LPs increases because of the *opportunity cost* effect. Lemma 2 thus shows that the sum of these two effects reverse the ranking between the minimum required rates of return.

Lemma 2 implies that the cost of capital for GPs in a H-fund necessarily exceeds that of a L-fund. This implies that the only rationale left for GPs to raise H-funds is to avoid the investment distortion when raising a L-fund. However, this distortion needs not arise with a liquid secondary market. When the secondary market price is high enough, the GP can call enough capital in period 1, as shown by equation (30). Second, when this investment distortion cost exists in L-funds, it arises precisely when the discount to NAV in the secondary market is large (low price P_L). This implies that the extra cost of capital for H-funds is large by Lemma 2. In fact, we will show that GPs may never raise H-funds when claims trade at a discount.

Assumption 6 (No fund segmentation)

$$\left[1 + r(\lambda, \delta_H)\right] \left[1 - \lambda + \frac{\lambda \delta_H (p + (1-\lambda)q)}{(1-\lambda)q}\right] \geq pR$$

We will show that under Assumption 6, the return H-LPs can make in the secondary market when claims trade at a discount is so high that GPs cannot offer a return acceptable to H-LPs in the primary market.[29] We may now state the main result of this section.

[29]We want to show that the *opportunity cost effect* can be so strong that GPs may stop raising H-funds when secondary claims trade at a discount. In Appendix C, we also show, however, that this result is not generic. When λ is low enough, H-LPs also invest in the primary market in funds that offer higher returns than L-funds. Hence, the primary market segmentation result of Proposition 2 is weaker with a secondary market but it may survive. We discuss this point further at the end of the section.

Proposition 3 (Secondary Market Equilibrium)

There are thresholds $0 < \mu_{H,1} < \mu_{H,2} < \mu_{H,3} < \bar{\mu}_H$ *for the share of* H*-LPs capital such that*

1) For $\mu_H \leq \mu_{H,1}$, $r^*_{PE,L} = r(\lambda, \delta_L)$, $D^* = 1 - \delta_L$, $x^*_L = \hat{x}(\lambda, \delta_L)$ *and* $\alpha^*_H = 0$

2) For $\mu_H \in [\mu_{H,1}, \mu_{H,2}]$, $x^*_L = 1 - P^*_L > \bar{x}$, $\alpha^*_H = 0$ *and* $(r^*_{PE,L}, P^*_L)$ *solve*

$$r^*_{PE} = \underline{r}_{PE,L} \tag{38}$$

$$\left[I\left(r^*_{PE}, x^*_L\right) - A\right] P^*_L = \mu_H M \tag{39}$$

3) For $\mu_H \in [\mu_{H,2}, \mu_{H,3}]$, $x^*_L = \bar{x}$, $\alpha^*_H = 0$ *and* (r^*_{PE}, P^*_L) *solve* (38)-(39).

4) For $\mu_H \geq \mu_{H,3}$, $r^*_{PE,L} = r^*_{PE,H} = r(\lambda, \delta_H)$, $D^* = 1 - \delta_H$, $x^*_L = \bar{x}$.

The equilibrium allocation is pinned down up to the primary market allocation when the share of H-LPs' capital is high enough ($\mu_H \geq \mu_{H,3}$). In this case, H-LPs and L-LPs are identical investors from the point of view of GPs. The L-LPs require the same return $r(\lambda, \delta_H)$ on their capital, equal to the break-even rate of H-LPs, because L-LPs can sell their claim in the secondary market at a price reflecting the value of the claim to an H-LP buyer. This also implies that GPs need not distort the investment schedule in a L-fund and thus that H-funds and L-funds are identical. With a secondary market, the parameter region $[\mu_{H,3}, 1]$ where GPs can raise capital at the minimal cost $r(\lambda, \delta_H)$ is larger since $\mu_{H,3} < \bar{\mu}_H$. The intuition is that L-LPs can now make the primary market thanks to the liquidity provided by H-LPs in the secondary market. The scarce resources of H-LPs are redeployed towards their more efficient use in the secondary market. We can show that the introduction of a secondary market lowers the cost of capital for GPs for all parameter values.

When H-LPs capital is less abundant, that is when $\mu_H \leq \mu_{H,3}$, there is cash-in-the-market pricing and the equilibrium discount satisfies $D^* > 1 - \delta_H$. In this case, H-LPs do not have enough resources collectively to bid the price up to their reservation value. In other words, they receive a compensation for providing liquidity in the secondary market.

By Lemma 2, H-LPs then require a strictly higher rate of return than L-LPs in the primary market. Hence, GPs only raise L-funds while H-LPs allocate all their resources to the secondary market. When $\mu_H \in [\mu_{H,2}, \mu_{H,3}]$, the discount is still small and the secondary market price lies above the default threshold $1 - \bar{x}$ for L-LPs. Hence, in a L-fund, GPs can still implement an optimal investment schedule. GPs do not raise H-funds because H-LPs require a higher cost of capital and the investment schedule is already optimal in a L-fund.

When $\mu_H \leq \mu_{H,2}$, liquidity is very scarce and the secondary market price is low. As a result, GPs cannot implement an optimal investment schedule in L-funds because L-LPs would rather default on the second capital call. Instead, as in the case without a secondary market, GPs increase the share of the fund capital x_L^* called in period 0. When the fraction of H-LPs investors decreases, liquidity in the secondary market decreases and the secondary market price goes down. This decreases the GPs' profit through two channels: L-LPs require a higher return $r_{PE,L}^*$ in the primary market to compensate for the lower liquidity of the secondary market and the investment schedule is distorted to avoid LPs default.

When the share of H-LPs capital goes below the threshold $\mu_{H,1}$, the secondary market price attains its minimum level equal to the reservation value of L-LPs. Then, in the region $[0, \mu_{H,1}]$, L-LPs are indifferent between selling or holding their claim to maturity. Hence, by definition, the cost of capital for L-funds is equal to the break-even rate $r(\lambda, \delta_L)$ of L-LPs. The profit of GPs in a L-fund is the same as in the case without a secondary market. Why do GPs not raise H-funds then? This is because H-LPs capture large discounts in the secondary market when the liquidity shock hits. Because of the *opportunity cost effect*, they require a very high return to invest in the primary market. Under Assumption 6, this required return is so high that GPs prefer not to raise H-funds although they could implement an optimal investment schedule such that $x_H^* = \bar{x}$.

To summarize this first set of results, the secondary market generates additional competition on the investor side. Interestingly, H-LPs facilitate the competition from L-LPs in the primary market by providing liquidity in the secondary market. This competition lowers H-LPs returns. For example, in the region where the share of H-LPs' capital satisfies $\mu_H \in [\mu_{H,3}, \bar{\mu}_H]$, H-LPs only make a profit when there is no secondary market. Hence, while it is privately optimal for a H-LP to supply liquidity in the secondary market, these investors would collectively benefit from shutting down the secondary market. GPs on the other hand benefit from the secondary market thanks to the lower cost of capital brought about by the *liquidity effect*.

6.4 Returns persistence with a secondary market

Under Assumption 6, we show that essentially one type of fund is offered in equilibrium. This contrast with the segmentation result in Proposition 2 where in the absence of a secondary market, two types of funds with different returns necessarily coexist for μ_H low enough. Our model thus suggests that the presence of a secondary market tends to reduce segmentation in the primary market (see also footnote 29). Hence, part of the recent decrease in fund-level persistence documented by Harris et al. (2014b) may be due to the growth of the secondary market for LP partnership claims. Consistently with this prediction, these authors show that persistence is more robust for VC funds where the secondary market is less mature.

Our model does not imply that return persistence should also disappear at the LP level when premium capital from H-LPs is scarce. Accounting for secondary market investments, H-LPs still realize higher returns on their private equity portfolio than L-LPs. The average monetary return on a dollar committed to a L-fund for a L-LP is given by

$$1 + r^{avg}_{PE,L} := (1 - \lambda)(1 + r_{PE,L}) + \lambda P_L = 1 \tag{40}$$

where the second equality is implied by equations (32) and (33) and the fact that a L-LP always breaks even. The average monetary return on a dollar committed to the secondary market by a H-LP is equal to

$$1 + r^{avg}_{c,H} := \lambda + (1 - \lambda)\frac{1 + r_{PE,L}}{P_L} > 1 \tag{41}$$

The inequality obtains because $P_L \leq \delta_H(1 + r_{PE,L})$ in equilibrium. Hence, H-LPs still earn a higher monetary return than L-LPs because they focus on the secondary market.

Interestingly, an H-LP would also perform better even if he could only invest in the primary market, that is $r^{avg}_{PE,H} > r^{avg}_{PE,L}$ with the notation introduced above. Their average monetary return from investing in the L-funds offered in equilibrium is given by

$$1 + r^{avg}_{PE,H} = 1 + r_{PE,L} > 1 \tag{42}$$

The difference between equations (40) and (42) follows from the fact that unlike a L-LP, a H-LP would not sell his claim when hit by a liquidity shock. Since secondary claims trade at a discount, the observed return is higher for a H-LP.

Finally, note that in the main text we focused for simplicity on a case without primary market segmentation in equilibrium (Assumption 6). As we show in Appendix C, under alternative parameter configurations, segmentation is still an equilibrium outcome. Then, the source of return differences between LPs would be the same than in Section 5 since H-funds would offer higher returns than L-funds.

7 Conclusions

This paper provides a model of delegated investment in private equity funds where investors are subject to liquidity risk. We derive the optimal partnership between GPs and LPs with a fund structure and a compensation contract that resemble actual partnership agreements. Because investors face liquidity risk, there is a pecking order for LPs' capital. GPs prefer to raise capital from LPs who are less sensitive to liquidity risk. These good LPs supply capital at a lower cost and are more likely to stand by their capital commitment. This last feature implies that when high-quality capital is scarce, GPs pay a premium to good LPs. This finding rationalizes persistent differences in returns between PE investors. Our model thereby suggests that GPs cherry-pick their investors for their ability to provide long-term capital. We also study the introduction of a secondary market for LP claims. Good LPs migrate from the primary market to the secondary market. Discounts to NAV arise endogenously when equilibrium liquidity is scarce. Finally, our model suggests that fund-level persistence may disappear with a secondary market while LP-level persistence remains.

Although our model is static, the insights can be useful to understand some private equity market dynamics. It is a well documented fact that private equity fundraising and dealmaking are highly procyclical. Our model identifies two potential supply-side explanations. In downturns, investors might require a higher illiquidity premium and GPs' wealth is low. Both these effects contribute to a decrease in fund size and a lower level of fundraising.

Finally, our analysis rests solely on two factors: the agency problem between fund managers and investors and the investors heterogeneous exposure to liquidity shocks. Our model does not exhibit investor irrationality, or asymmetric information and/or learning about GP skills but we do not necessarily believe such features are not important in practice. Instead, we provide a benchmark against which one can judge whether the observed patterns are consistent with agents being informed and rational. In this sense, we offer a counterpart to the Berk and Green (2004) model of mutual fund for investments in liquid assets. We believe the stylized structure of our model makes it applicable to delegated portfolio management in other illiquid asset classes, such as infrastructure, private credit, or real estate funds.

Appendix

A Proofs

A.1 Proof of Proposition 1

Building on our analysis in the main text, we are left to show two results. The first is that equation (11) holds, that is the GP is only compensated after a joint success. The second result is that it is suboptimal for the GP to increase the promised return over $r(\lambda, \delta)$ in order to relax the no-default constraint (5).

Proof that equation (11) holds

As stated in the text, since the GP is risk-neutral, he should only be compensated after the outcome most informative about effort exertion. For each relevant unit payoff y in (4), we can then define an informativeness ratio

$$\mathcal{I}(y) = \frac{\mathbf{Pr}[\tilde{y} = y|\text{effort}]}{\mathbf{Pr}[\tilde{y} = y|\text{shirk}]} \tag{A.1}$$

The higher $\mathcal{I}(y)$, the better signal of effort is payoff y. Two cases need to be considered: either the GP shirks on both investments or he only shirks on the first investment. By symmetry, the argument is similar if he only shirks on the second investment. If the GP shirks on both investments, the probability of a joint success if q^2. We thus have

$$\mathcal{I}(R) = \frac{p^2}{q^2} > \frac{p(1-p)}{q(1-q)} = \mathcal{I}(Rx) = \mathcal{I}(R(1-x))$$

When the GP shirks on both investments, a single success obtains with probability $q(1-q)$. The strict inequality follows from $p > q$. Suppose now that the GP only shirks on the first investment. This means that the probability of a joint success when shirking is pq. The probability of a success of the first investment only is $q(1-p)$ and the probability of a success of the second investment only is $p(1-q)$. We now have

$$\mathcal{I}(R) = \frac{p^2}{pq} = \frac{p(1-p)}{q(1-p)} = \mathcal{I}(Rx) > \frac{p(1-p)}{q(1-q)} = \mathcal{I}(R(1-x))$$

This shows that is always weakly optimal to compensate the GP only in the state $y = R$ and that equation (11) holds.

Proof that $r^*_{PE} = r(\lambda, \delta)$

To verify our claim in the main text, we have to prove that when $\delta < \hat{\delta}(\lambda)$, a GP never finds it optimal to decrease x from $\hat{x}(\lambda, \delta)$ by offering a return strictly above $r(\lambda, \delta)$. Under the binding no-default constraint (5), the expected return to the LPs would be given by $\hat{r}_{PE}(x) = (1 - x)/\delta$ where $x \in [\bar{x}, \hat{x}(\lambda, \delta)]$ is chosen by the GP. The expected compensation of the GP is given by $\underline{W}_x = \frac{pBx}{p-q}$. The profit of the GP as a function of x is this alternative fund is given by

$$\hat{\Pi}(x) = \frac{A\underline{W}_x}{1 - \frac{\delta}{1-x}\left(pR - \underline{W}_x\right)}$$

We are left to show that $\hat{\Pi}(\hat{x}(\delta)) > \hat{\Pi}(\bar{x})$. Since the numerator is increasing in x, it is enough to show that the denominator is decreasing in x. We thus have

$$\begin{aligned} 0 \leq \frac{\partial \hat{\Pi}}{\partial x} \quad &\Leftarrow \quad 0 \leq -\frac{\delta}{1-x}\frac{pB}{p-q} + \frac{\delta}{(1-x)^2}(pR - \frac{pBx}{p-q}) \\ &\Leftarrow \quad 0 \leq \frac{\delta}{(1-x)^2}\left[pR - \frac{pB}{p-q}\right] \end{aligned}$$

The last inequality follows from Assumption 1. This proves that $\frac{\partial \hat{\Pi}}{\partial x} \geq 0$ and that GPs do not increase the promised return to relax the no-default constraint (5).

A.2 Proof of Corollary 1

– *Effect of* $1 - \delta$

This result follows directly from the discussion in the text.

– *Effect of* λ

When $\delta \geq \hat{\delta}(\lambda)$, an increase in λ increases the required rate of return $r(\lambda, \delta)$. This lowers fund size and profit. We now turn to the case $\delta \leq \hat{\delta}(\lambda)$. In order to use our previous results, let us then write Π^*_{GP} and I^* as a function of x. We have

$$I^* = \frac{A}{1 - \frac{\delta}{1-x}\left(pR - \frac{pBx}{p-q}\right)}, \qquad \Pi^*_{GP} = I^* \frac{pBx}{p-q}$$

We showed in the Proof of Proposition 1 that I^* and Π^*_{GP} are increasing in x. When $\delta \leq \hat{\delta}(\lambda)$, by equation (18), we have that $x = \hat{x}(\lambda, \delta)$ is a decreasing function of λ. This proves that profit and fund size are decreasing in λ.

– *Effect of R*

This result follows immediately from the inspection of equation (20) and (21).

– *Effect of q when $p = q + \alpha$ with α constant*

Expression (20) shows that the fund size I^* depends on q only through the pledgeable income

$$pR - W^* = (q+\alpha)R - \begin{cases} \frac{(q+\alpha)^2}{\alpha(2q+\alpha)}B & \text{if} \quad \delta \geq \hat{\delta}(\lambda) \\ \frac{q+\alpha}{\alpha}B\hat{x}(\lambda,\delta) & \text{if} \quad \delta < \hat{\delta}(\lambda) \end{cases}$$

The result when $\delta < \hat{\delta}(\lambda)$ follows directly from the fact that $R \geq B/\alpha$ by Assumption 1. In the case where $\delta \geq \hat{\delta}(\lambda)$, a similar conclusion arises since $\frac{\partial(q+\alpha)}{\partial q} > 0$ and $\frac{\partial\left(\frac{q+\alpha}{2q+\alpha}\right)}{\partial q} < 0$ The profit Π^*_{GP} is equal to the expected fee W^* multiplied by the fund size I^*. The expression above shows that the expected fee is increasing in q. Since we showed that the fund size is also increasing in q, this proves that the profit of the GP is increasing in q.

A.3 Proof of Proposition 2

We first prove a preliminary result before analyzing the three cases of Proposition 2.

Proof that if $\alpha^*_L > 0$, then $r^*_{PE,L} = r(\lambda, \delta_L)$

That $r^*_{PE,L} \geq r(\lambda, \delta_L)$ is obvious since $r(\lambda, \delta_L)$ is the minimum rate of return required by L-LPs. To prove the reverse inequality, we proceed by contradiction. If $r^*_{PE,L} > r(\lambda, \delta_L)$, L-LPs invest all their resources in PE funds, that is $S^L = \mu_L M$. Let us now prove that H-LPs would also invest all their resources in PE funds to arrive at a contradiction. Observe first that it must be that $\alpha^*_H > 0$ when $\mu_H > 0$. If $\alpha^*_H = 0$, from equation (24) and by market clearing it must be that $r^*_{PE,H} \leq r(\lambda, \delta_L)$. But GPs would then strictly prefer to raise funds from H-LPs since

$$\Pi^H(r^*_{PE,H}) \geq \Pi^H(r(\lambda, \delta_H)) > \Pi^L(r(\lambda, \delta_L)) > \Pi^L(r^*_{PE,L})$$

The first and the last inequality hold because Π^i is decreasing in $r_{PE,i}$. The middle inequality was proved in Section 1. Hence, this cannot be an equilibrium and it must be that $\alpha_H^* > 0$ when $\alpha_L^* > 0$. Finally, by optimality of GPs' decision, the return offered to H-LPs, denoted $r_{PE,H}^*$ would then satisfy

$$\Pi^H(r_{PE,H}^*) \geq \Pi^L(r(\lambda, \delta_L)) \tag{A.2}$$

so that in particular $r_{PE,H}^* > r(\lambda, \delta_H)$. But equation (24) then implies that $S^H = \mu_H M$. Hence, the total supply of capital by LPs is M while by assumption 3, the demand from GPs is strictly below M. This cannot be an equilibrium since markets would not clear. Our analysis also establishes that $r_{PE,H}^* \in [r(\lambda, \delta_H), \bar{r}_{PE,H}]$ where $\bar{r}_{PE,H}$ is the value of $r_{PE,H}$ that satisfies (A.2) as an equality. In addition, we must have $\alpha_L^* = 0$ when $r_{PE,H}^* < \bar{r}_{PE,H}$. We now examine the three possible cases.

Three different cases

Case i) Suppose first that $r_{PE,H}^* = r(\lambda, \delta_H)$. This implies that $\alpha_L^* = 0$. With the supply of capital by H-LPs given by the second case in equation (24), market clearing requires that $I_H(r(\lambda, \delta_H)) \leq \mu_H M + A$. This holds if $\mu_H \geq \bar{\mu}_H$ where

$$\bar{\mu}_H = \frac{I_H(r(\lambda, \delta_H)) - A}{M} \tag{A.3}$$

This proves the first case of Proposition 2.

Case ii) Suppose now that $r_{PE,H}^* \in (r(\lambda, \delta_H), \bar{r}_{PE,H})$. Once again, we have $\alpha_L^* = 0$. From (24), the supply of capital from H-LPs is given by $\mu_H M$ so market clearing requires that

$$\mu_H M + A = I_H(r_{PE,H}^*) \tag{A.4}$$

which implicitly defines $r_{PE,H}^*$ as as strictly decreasing function of μ_H. Comparing (A.3) and (A.4), the inequality $r_{PE,H}^* > r(\lambda, \delta_H)$ implies that this outcome can only be an equilibrium if $\mu_H \leq \bar{\mu}_H$. Similarly, the inequality $r_{PE,H}^* < \bar{r}_{PE,H}$ imposes a lower bound $\underline{\mu}_H$ on μ_H. Since $\bar{r}_{PE,H}$ is implicitly defined by the condition $\Pi^H(\bar{r}_{PE,H}) = \Pi^L(r(\lambda, \delta_L))$, using that $\Pi^H(r_{PE,H}) = W_H^* I_H(r_{PE,H})$ together with equation (A.4), this lower bound is given by

$$\underline{\mu}_H := \frac{\frac{\Pi^L(r(\lambda, \delta_L))}{W_H^*} - A}{M}$$

Re-arranging equation (A.4) and using the definition of $\bar{\mu}_H$ in (A.3),

$$1 + r^*_{PE,H} = (pR - W^*_H)\frac{A + \mu_H M}{A} = (1 + r(\lambda, \delta_H))\frac{I_H(r(\lambda, \delta_H)) - A}{I_H(r(\lambda, \delta_H))}\frac{A + \mu_H M}{A}$$
$$= \frac{\bar{\mu}_H(A + \mu_H M)}{\mu_H(A + \bar{\mu}_H M)}\big(1 + r(\lambda, \delta_H)\big)$$

which proves the expression in Case 2 of Proposition 2.

Case iii) Finally, consider the case where $r^*_{PE,H} = \bar{r}_{PE,H}$. In this case, the capital supply from L-LPs is indeterminate since $r^*_{PE,L} = r(\lambda, \delta_L)$. The supply of capital from H-LPs is $S_H = \mu_H M$ since $r^*_{PE,H} > r(\lambda, \delta_H)$. Hence, market clearing for funds with H type investors requires that

$$\alpha^*_H(I_H(\bar{r}_{PE,H}) - A) = \mu_H M$$

which pins down α^*_H. This equation implies that α^*_H is an increasing function of μ_H over $[0, \underline{\mu}_H]$ with $\alpha^*_H(\underline{\mu}_H) = 1$.

Hence, we showed that the allocation in Proposition 2 is an equilibrium. Our analysis of the three cases also shows that this is the unique equilibrium.

A.4 Proof of Corollary 2

Using equation (26), we obtain

$$\frac{\partial \bar{r}_{PE,H}(R)}{\partial R} \propto W^*_L - W^*_H - \frac{W^*_L W^*_H}{1 + r(\lambda, \delta_L)} + \frac{W^*_H}{1 + r(\lambda, \delta_L)}W^*_H = (W^*_L - W^*_H)\left[1 - \frac{W^*_H}{1 + r(\lambda, \delta_L)}\right]$$

which proves our first result.

We now prove the matching result when $W^*_H < 1 + r(\lambda, \delta_L)$. We need to show that the equilibrium return for H-LPs is strictly above the threshold $\bar{r}_{PE,H}(R_b)$ when μ_H is low. Suppose by contradiction that $r^*_{PE,H} \leq \bar{r}_{PE,H}(R_b)$. Under the condition above, good GPs strictly prefer to raise funds from H-LPs since $\bar{r}_{PE,H}(R_g) > \bar{r}_{PE,H}(R_b)$. Their demand for capital is thus strictly bounded below by 0. But as $\mu_H \to 0$, the supply of capital from H-LPs converges to 0. This cannot be an equilibrium. Hence, when μ_H is too low, it must be that $r^*_{PE,H} > \bar{r}_{PE,H}(R_b)$ to clear the market. When this is the case, H-LPs only supply capital to good GPs.

A.5 Proof of Corollary 3

– *Effect of* λ

Observe that λ affects the premium through $r(\lambda, \delta_L)$ and W_L^*. Given that an increase in λ increases $r(\lambda, \delta_L)$ and decreases W_L^*, it follows that the premium is decreasing in λ since the premium is negatively affected by $r(\lambda, \delta_L)$ and positively affected by W_L^*.

– *Effect of* δ_L

We showed that W_L^* and $r(\lambda, \delta_L)$ are decreasing in δ_L. By the argument above, it follows that the effect of δ_L on the premium is ambiguous.

– *Effect of* R

Let us call rp the return premium. We have that

$$\begin{aligned}\frac{\partial rp}{\partial R} &\propto (W_L^* - W_H^*)(1 + r(\lambda, \delta_L)) + W_H^*(pR - W_L^*) - W_H^*(pR - W_H^*)) \\ &= (W_L^* - W_H^*)(1 + r(\lambda, \delta_L) - W_H^*)\end{aligned}$$

The result follows since $W_L^* > W_H^*$.

A.6 Proof of Lemma 2

Rewriting equations (32) and (35) using the definition of the discount in equation (37), we obtain

$$\begin{aligned}1 + u_{PE,L} &= (1 + r_{PE,L})\,[1 - \lambda + \lambda(1 - D)] \\ 1 + u_{c,H} &= 1 - \lambda + \lambda\delta_H \frac{1}{1 - D}\end{aligned}$$

By definition, $\underline{r}_{PE,L}(D)$ and $\underline{r}_{PE,H}(D)$ are respectively the values of $r_{PE,L}$ and $r_{PE,H}$ such that $u_{PE,L} = u_{c,L}$ and $u_{PE,H} = u_{c,H}$. We thus have

$$\underline{r}_{PE,L}(D) = \frac{\lambda D}{1 - \lambda D} \tag{A.5}$$

$$\underline{r}_{PE,H}(D) = \frac{\lambda\delta_H D}{(1 - D)\big[1 - \lambda(1 - \delta_H)\big]} \tag{A.6}$$

Subtracting these two equations, we obtain

$$\underline{r}_{PE,H}(D) - \underline{r}_{PE,L}(D) = \frac{\lambda D}{1-\lambda(1-\delta_H)}\left[\frac{\delta_H}{1-D} - \frac{1-\lambda(1-\delta_H)}{1-\lambda D}\right]$$
$$= \frac{(1-\lambda)\lambda D}{1-\lambda(1-\delta_H)}\frac{D-(1-\delta_H)}{(1-D)(1-\lambda D)}$$

Since the numerator is increasing in D and the denominator is decreasing, this proves that $\underline{r}_{PE,H}(D) - \underline{r}_{PE,L}(D)$ is increasing in D. The expression above also shows that the difference is equal to 0 when $D = 1-\delta_H$.

A.7 Proof of Proposition 3

The proof is in two steps. We first characterize the equilibrium under the conjecture that H-LPs do not participate in the primary market when $D > 1-\delta_H$. Then, we verify that it is optimal for GPs not to raise H-funds.

Equilibrium characterization

Following the proof of Proposition 2., we prove the result by construction. For each possible value of the discount D_*, we characterize the equilibrium and the range of values for μ_H where this equilibrium exists.

Case 1. $D^* = 1-\delta_H$.

From Lemma 2, we obtain that $\underline{r}_{PE,L} = \underline{r}_{PE,H} = r(\lambda, \delta_H)$.

Since

$$P_L^* = \delta_H(1+r_{PE,L}) > 1-\bar{x}$$

by Assumption (5), the capital call in a L-fund is given by $x_L = \bar{x}$. Hence, by optimality of the fund choice, it must be that $r^*_{PE,L} = r^*_{PE,H}$. By the clearing condition in the primary market, we further obtain that $r^*_{PE,L} = r(\lambda, \delta_H)$ since otherwise the supply of funds from LPs would exceed the demand by GPs. This allocation can be an equilibrium if and only if the supply of claims in the secondary market exceeds the supply at price $P_L^* = \delta_H(1+r(\lambda,\delta_H))$, that is

$$\big[I(r(\lambda,\delta_H),\bar{x}) - A - S_H\big]\delta_H\big[1+r(\lambda,\delta_H)\big] \leq \mu_H M - S_H \tag{A.7}$$

Note that because $\delta_H\big[1+r(\lambda,\delta_H)\big] < 1$, this inequality is easier to satisfy when $S_H = 0$. Hence, this allocation is an equilibrium for $\mu_H \geq \mu_{H,3}$ where $\mu_{H,3}$ is the minimum value of μ_H such that equation (A.7) holds when setting $S_H = 0$. We thus obtain

$$\mu_{H,3} = \frac{\left[I(r(\lambda, \delta_H), \bar{x}) - A\right]\delta_H\left[1 + r(\lambda, \delta_H)\right]}{M}$$

Again, because $\delta_H\left[1 + r(\lambda, \delta_H)\right] < 1$, a comparison between the equation above and equation (A.3) shows that $\mu_{H,3} < \bar{\mu}$.

Case 2: $D^* = 1 - \delta_L$.

This implies that $\underline{r}_{PE,L} = r(\lambda, \delta_L)$. Since GPs only raise L-funds, the clearing condition in the primary market implies that $r^*_{PE,L} = r(\lambda, \delta_L)$. Combining this result with $D = 1 - \delta_L$, we obtain $x^* = \hat{x}(\lambda, \delta_L)$. By equation (27), the supply of claims in the secondary market is indeterminate. This outcome is an equilibrium if the maximum supply of claims in the secondary market exceeds the demand at price $P^*_L = \delta_L(1 + r(\lambda, \delta_L))$. Using equation (28), the condition writes

$$\left[I(r(\lambda, \delta_L), \hat{x}(\lambda, \delta_L)) - A\right]\delta_L(1 + r(\lambda, \delta_L)) \geq \mu_H M \tag{A.8}$$

which we rewrite as $\mu_H \leq \mu_{H,1}$. Since $I(r, x)$ is decreasing in r and x for $x \geq \bar{x}$ and $\delta(1 + r(\lambda, \delta))$ is increasing in δ, the comparison between equations (A.7) and (A.8) shows that $\mu_{H,1} < \mu_{H,3}$.

Case 3: $D^* \in (1 - \delta_H, 1 - \delta_L)$.

Since $P^*_L > \delta_L(1 + r^*_{PE,L})$, L-LPs strictly prefer to sell their claims by equation (27). Since $P^*_L < \delta_H(1 + r^*_{PE,L})$, the demand for claims from H-LPs is given by $\mu_H M$. Hence, the market clearing condition on the secondary market writes

$$\left[I(r^*_{PE,L}, x^*_L) - A\right]P^*_L = \mu_H M \tag{A.9}$$

The binding participation constraint of L-LPs implies that

$$r^*_{PE,L} = \underline{r}_{PE,L} = \frac{1 - \lambda P^*_L}{1 - \lambda} \tag{A.10}$$

By definition, $I(r, x)$ is strictly decreasing in r and weakly decreasing in x (strictly over the range $[\bar{x}, \hat{x}(\lambda, \delta_L)]$). Since $r^*_{PE,L}$ and x^*_L are themselves strictly and weakly decreasing in P^* respectively.

the left hand side of (A.9) is strictly increasing in P_L^* and thus decreasing in D^*. Hence, condition $D^* \in (1-\delta_H, 1-\delta_L)$ defines a range of values of μ_H where equation (A.9) may hold. Comparing equations (A.9) to equations (A.7) and (A.8), the upper bound and lower bound of this region are given respectively by $\mu_{H,3}$ and $\mu_{H,1}$. Over this region, P^* and r_{PE}^* are strictly monotone in μ_H.

To finish the equilibrium characterization, let us define $\mu_{H,2}$ as the value of $\mu_H \in [\mu_{H,1}, \mu_{H,3}]$ such that equation (A.9) holds with $x_L^* = \bar{x}$, $P_L^* = 1 - \bar{x}$ and where $r_{PE,L}$ is given by equation (A.10). Since we showed that P_L^* is increasing over $[\mu_{H,1}, \mu_{H,3}]$, this implies that $x_L^* = \bar{x}$ for $\mu_H \in [\mu_{H,2}, \mu_{H,3}]$ using equation (30) for the optimal choice of x_L by GPs..

Note that the parameter values for an equilibrium with two different values of D^* are mutually exclusive. This implies that we characterized the only equilibrium that satisfies our conjecture. We now verify this conjecture that H-LPs do not invest in the primary market when $D^* > 1 - \delta_H$, that is when $\mu_H < \mu_{H,3}$.

No fund with H-LPs when $\mu_H \leq \mu_{H,3}$

To prove this result, we show that the minimum return $\underline{r}_{PE,H}(D^*)$ required by H-LPs exceeds the maximum return $\bar{r}_{PE,H}$ that GPs are willing to pay. This return is the value of $r_{PE,H}$ such that GPs are indifferent between H-funds and L-funds. Adapting the equation we already derived in (26), we obtain

$$1 + \bar{r}_{PE,H} = \frac{\underline{W}_{x_L^*}(pR - \underline{W}_{\bar{x}})}{\underline{W}_{\bar{x}}(pR - \underline{W}_{x_L^*}) + (\underline{W}_{x_L^*} - \underline{W}_{\bar{x}})(1 + r_{PE,L}^*)}(1 + r_{PE,L}^*) \tag{A.11}$$

When $\mu_H \in [\mu_{H,2}, \mu_{H,3}]$, we showed that $x_L^* = \bar{x}$. Hence, equation (A.11) becomes $\bar{r}_{PE,H} = r_{PE,L}^*$. Since $D^* > 1 - \delta_H$, the minimum rate required by H-LPs satisfies $\underline{r}_{PE,H}(D^*) > r_{PE,L}^*$ by Lemma 2. This proves our claim in this case.

For $\mu_H \leq \mu_{H,2}$, we write $\bar{r}_{PE,H}$ as a function of x_L^* using equation (30) and (A.10) to substitute for $r_{PE,L}^*$ in (A.11). We obtain

$$1 + \bar{r}_{PE,H} = \frac{\underline{W}_{x_L^*}(pR - \underline{W}_{\bar{x}})\left[1 - \lambda(1 - x_L^*)\right]}{(1-\lambda)\underline{W}_{\bar{x}}(pR - \underline{W}_{x_L^*}) + (\underline{W}_{x_L^*} - \underline{W}_{\bar{x}})\left[1 - \lambda(1 - x_L^*)\right]} \tag{A.12}$$

Since $\bar{r}_{PE,H}$ is increasing in $\underline{W}_{x_L^*}$ and $r_{PE,L}^*$ which are themselves increasing in x_L^*, we obtain that $\bar{r}_{PE,H}$ is increasing in x_L^*. We can similarly write $\underline{r}_{PE,H}$ as a function of x_L^* using equation (A.6) and writing D^* as a function of x_L^*. We obtain

$$1 + \underline{r}_{PE,H} = 1 + \frac{\lambda \delta_H x_L^*}{(1-\lambda)(1-x_L^*)\big[1-\lambda(1-\delta_H)\big]} \tag{A.13}$$

which is also increasing in x_L^*. We showed that x_L^* is strictly decreasing in μ_H for $\mu_H \in [\mu_{H,1}, \mu_{H,2}]$. For $\mu_H \leq \mu_{H,1}$, x_L^* is constant and equal to $\hat{x}(\lambda, \delta_L)$. Hence, in order to prove that $\underline{r}_{PE,H} > \bar{r}_{PE,H}$ for all values of $\mu_H \in [0, \mu_{H,2}]$, it is enough to show that $\underline{r}_{PE,H}(x = \bar{x}) \geq \bar{r}_{PE,H}(x = \hat{x}(\lambda, \delta_L))$. An upper bound on $\bar{r}_{PE,H}(x = \hat{x}(\lambda, \delta_L))$ is given by $pR - 1$. On the other hand, we have

$$\underline{r}_{PE,H}(x = \bar{x}) = \frac{\lambda \delta_H p}{(1-\lambda)q\big[1-\lambda(1-\delta_H)\big]}$$

Hence, under Assumption 6, the result obtains. This concludes the proof.

B Contract Robustness

B.1 Calling Excess Capital

We show that GPs cannot avoid the commitment problem of LPs by calling excess capital in period 0. The key intuition for the result is that GPs would deviate by choosing to invest in period 0 all the capital called. We assume that LPs observe the realized investment before the GP chooses the effort level and that

$$R \geq \frac{B}{(p-q)^2} \tag{B.1}$$

This last assumption will ensure that LPs cannot commit to punish the GP by seizing all the cash flows. Inequality (B.1) is compatible with assumptions 1 and 2 in the main text. In our analysis, we focus on the case $\delta < \hat{\delta}(\lambda)$ since otherwise, Proposition 1 shows that the commitment problem of LPs is moot.

Claim B.1. *When $\delta < \hat{\delta}(\lambda)$, GPs cannot increase profit by holding LPs' capital as cash.*

Proof. To avoid default by LPs, the GP must call at least a fraction $\hat{x}(\lambda, \delta)$ of the fund capital in period 0 where $\hat{x}(\lambda, \delta)$ is defined in Proposition 1. Without loss of generality, we assume that the GP calls exactly $\hat{x}(\lambda, \delta)$. We denote by $x_{inv} \in \left[\frac{p}{p+q}, \hat{x}(\lambda, \delta)\right]$ the amount that the GP should invest in period 0 according to the partnership contract. The lower bound on x_{inv} is without loss of generality since Proposition 1 shows that the diversification benefits are maximized for this value. Given the fund size I, the GP expected compensation if he invests according to the contractual schedule is given by

$$\Pi_{GP} = \frac{pBx_{inv}}{p-q} I$$

The GP can deviate by investing all the capital $\hat{x}(\lambda, \delta)I$ he is entrusted with in period 0. To punish the GPs, the LPs can pull out the second capital call. However, we conjecture and verify that LPs would still compensate the GP to induce effort on the first investment. In this case, the minimum wage compatible with effort on the first investment is $\tilde{w} = pB/(p-q)$ per unit of investment in case of success. The profit of the GP is then equal to

$$\tilde{\Pi}_{GP} = \frac{pB}{p-q} \hat{x}(\lambda, \delta) I > \Pi_{GP}$$

so that the GP will indeed find it optimal to deviate. The payoff to LPs from the first investment is

$$\tilde{\Pi}_{LP} = p(R - \tilde{w})\hat{x}(\lambda, \delta)I = p\left(R - \frac{B}{p-q}\right)\hat{x}(\lambda, \delta)I$$

We are left to check that the LPs would not try to punish the LPs following a deviation. Suppose that LPs confiscate the proceeds from the first investment. The GP would react by shirking, leaving a profit equal to $q\hat{x}(\lambda, \delta)RI$ to the LPs on the first investment. Under Assumption (B.1), this payoff is lower than $\tilde{\Pi}_{LP}$ derived above. Hence, the LPs cannot commit to punishing the GP after a deviation. This proves that the proposed contract is not incentive-compatible. Given that the GP would always deviate by investing all the capital $x(\lambda, \delta)I$ called in period 0, the contract might as well specify that $x_{inv} = x(\lambda, \delta)$, which is the contract considered in the main text. This concludes the proof.

□

B.2 Raising New Capital

We show that ex-post capital raised in period 1 is more expensive. Let us define $\tilde{r}_1$ as the cost of ex-post capital raised in period 1 where

$$1 + \tilde{r}_1 = \begin{cases} \frac{1}{\delta} & (\text{prob } \lambda) \\ 1 & (\text{prob } 1 - \lambda) \end{cases}$$

Claim B.2. *Ex-ante capital is cheaper than ex-post capital , that is* $r(\lambda, \delta) < \mathbb{E}[\tilde{r}_1]$.

Proof. We showed in the main text that

$$1 + r(\lambda, \delta) = \frac{1}{1 - \lambda + \lambda\delta} = \frac{1}{\mathbb{E}[\tilde{\delta}]}, \quad \text{where} \quad \tilde{\delta} = \begin{cases} \delta & (\text{prob } \lambda) \\ 1 & (\text{prob } 1 - \lambda) \end{cases}$$

With the notation above we have

$$\mathbb{E}[1 + \tilde{r}_1] = \lambda\frac{1}{\delta} + 1 - \lambda = \mathbb{E}\left[\frac{1}{\tilde{\delta}}\right] > \frac{1}{\mathbb{E}[\tilde{\delta}]} = 1 + r(\lambda, \delta)$$

where the result follows from Jensen's inequality applied to the convex function $x \mapsto 1/x$. □

C Fund Segmentation with a Secondary Market

We prove the claim in footnote 29 that essentially different H-funds and L-funds may coexist in the primary market when λ is low enough.

Claim C.1. *There exists $\hat{\lambda} > 0$ and $\epsilon > 0$ such that when $\lambda \leq \hat{\lambda}$ and $\delta_H - \delta_L < \epsilon$, both L-funds and H-funds are offered in equilibrium with $r^*_{PE,H} > r^*_{PE,L}$ for $\mu_H \leq \mu_{H,1}$.*

Proof. We proceed by contradiction and first show that the allocation in Proposition 3 cannot be an equilibrium under the conditions of Claim C.1. In particular, the fund optimality condition and the market clearing condition of Definition 2 are inconsistent. Given our analysis in the main text, this is equivalent to showing that the maximum rate GPs are willing to pay for H-funds $\bar{r}_{PE,H}$ exceeds the minimum rate $\underline{r}_{PE,H}$ H-LPs are willing to accept.

When $\mu_H \leq \mu_{H,1}$, the first capital call in a L-fund is $x_L^* = \hat{x}(\lambda, \delta_L)$ in the conjectured equilibrium. By continuity, it is enough to prove the result for $\delta_H = \hat{\delta}(\lambda)$. Using equation (A.13), we obtain

$$1 + \underline{r}_{PE,H} = (1 + r(\lambda, \hat{\delta}(\lambda)) \left[1 - \lambda + \lambda \hat{\delta}(\lambda) \frac{1 - \lambda(1 - \hat{x}(\lambda, \delta_L))}{(1 - \lambda)(1 - \hat{x}(\lambda, \delta_L))} \right]$$

Taking further the limit when $\delta_L \to \hat{\delta}(\lambda)$, we have

$$\lim_{\delta_L \to \hat{\delta}(\lambda)} (1 + \underline{r}_{PE,H}) = 1 + r(\lambda, \hat{\delta}(\lambda)) = \lim_{\delta_L \to \hat{\delta}(\lambda)} (1 + \bar{r}_{PE,H})$$

where the second equality follows from (A.12). Hence, to show the desired result and since $\hat{x}(\lambda, \delta_L)$ is decreasing in δ_L, it is enough to show that

$$\frac{\partial(1 + \underline{r}_{PE,H})}{\partial x}_{x = \hat{x}(\lambda, \hat{\delta}(\lambda))} < \frac{\partial(1 + \bar{r}_{PE,H})}{\partial x}_{x = \hat{x}(\lambda, \hat{\delta}(\lambda))} \tag{C.1}$$

We obtain

$$\frac{\partial(1 + \underline{r}_{PE,H})}{\partial x}_{x = \hat{x}(\lambda, \hat{\delta}(\lambda))} = \frac{(1 + r(\lambda, \hat{\delta}(\lambda)) \lambda \hat{\delta}(\lambda)}{(1 - \lambda) \left[1 - \hat{x}(\lambda, \hat{\delta}(\lambda)) \right]^2} = \frac{\lambda(p + q)}{(1 - \lambda) q} \tag{C.2}$$

For the right hand side of (C.1), using equation (A.12), we have

$$\begin{aligned}
\frac{\partial(1 + \bar{r}_{PE,H})}{\partial x} &= \frac{pB}{p - q} \frac{\partial \bar{r}_{PE,H}}{\partial \underline{W}_x} + \frac{\partial \bar{r}_{PE,H}}{\partial x} \\
&= \frac{pB}{p - q} \frac{(pR - \underline{W}_{\bar{x}})(1 - \lambda(1 - x)) \underline{W}_{\bar{x}} \left[pR(1 - \lambda) - 1 + \lambda(1 - x) \right]}{\left[(1 - \lambda) \underline{W}_{\bar{x}} \left(pR - \underline{W}_x \right) + (\underline{W}_x - \underline{W}_{\bar{x}})(1 - \lambda(1 - x)) \right]^2} \\
&\quad + \frac{\lambda(1 - \lambda) \underline{W}_x (pR - \underline{W}_{\bar{x}}) \underline{W}_{\bar{x}} (pR - \underline{W}_x)}{\left[(1 - \lambda) \underline{W}_{\bar{x}} \left(pR - \underline{W}_x \right) + (\underline{W}_x - \underline{W}_{\bar{x}})(1 - \lambda(1 - x)) \right]^2} \\
&= \frac{pB}{p - q} \underline{W}_{\bar{x}} (pR - \underline{W}_{\bar{x}}) \frac{(1 - \lambda(1 - x)) \left[pR(1 - \lambda) - 1 + \lambda(1 - x) \right] + x \lambda (1 - \lambda)(pR - \underline{W}_x)}{\left[(1 - \lambda) \underline{W}_{\bar{x}} \left(pR - \underline{W}_x \right) + (\underline{W}_x - \underline{W}_{\bar{x}})(1 - \lambda(1 - x)) \right]^2}
\end{aligned}$$

Setting $x = \hat{x}(\lambda, \hat{\delta}(\lambda)) = \bar{x}$ in the expression above, we obtain

$$\frac{\partial(1 + \bar{r}_{PE,H})}{\partial x}_{|x = \hat{x}(\lambda, \hat{\delta}(\lambda))} = \frac{\lambda}{1 - \lambda} + \frac{(1 - \lambda(1 - \bar{x}) \left[pR - \frac{1 - \lambda(1 - \bar{x})}{1 - \lambda} \right]}{(1 - \lambda) \bar{x} \left(pR - \underline{W}_{\bar{x}} \right)} \tag{C.3}$$

Note that the second term of (C.3) does not converge to 0 as $\lambda \to 0$. Hence, comparing (C.3) and (C.2), it follows that the required condition (C.1) holds when δ_H and δ_L are close enough to $\hat{\delta}$ and

λ is small enough. This implies that the allocation of Proposition 3 cannot be an equilibrium

The last step is to show that H-funds deliver higher return than L-funds in equilibrium. According to Lemma 2, this is true if the equilibrium discount is strictly higher than $1 - \delta_H$. Proposition showed that an equilibrium with a discount $D^* = 1 - \delta_H$ can only exist if $\mu_H \leq \mu_{H,3}$ where $\mu_{H,3} > \mu_{H,1}$. Hence, under the parameter configuration of Claim C.1, the discount is strictly higher than $1 - \delta_H$ which proves the claim. □

References

Albuquerque, Rui A., Johan Cassel, Ludovic Phalipou, and Enrique J. Schroth, 2018, Liquidity provision in the secondary market for private equity fund stakes, Working paper, Cass Business School.

Allen, Franklin, and Douglas Gale, 2005, From cash-in-the-market pricing to financial fragility, *Journal of the European Economic Association* 3, 535–546.

Axelson, Ulf, Per Strömberg, and Michael S. Weisbach, 2009, Why are buyouts levered? the financial structure of private equity funds, *The Journal of Finance* 64, 1549–1582.

Banat-Estañol, Albert, Filippo Ippolito, and Sergio Vicente, 2017, Default penalties in private equity partnerships, Working paper.

Berk, Jonathan, and Richard Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Bollen, Nicolas, and Berk Sensoy, 2016, How much for a haircut? illiquidity, secondary markets, and the value of private equity, Working paper, Fisher College of Business, Ohio State University.

Brown, Gregory, Robert Harris, Wendy Hu, Tim Jenkinson, Steven Kaplan, and Robinson Steve, 2018, Can investors time their exposure to private equity, Working paper, Private Equity Research Council.

Cagnavaro, Daniel, Berk Sensoy, Yingdi Wang, and Michael Weisbach, 2018, Measuring institutional investors' skill at making private equity investments, Working paper, Ohio State University.

Da Rin, Marco, Thomas Hellmann, and Manju Puri, 2012, A survey of venture capital research, in George Constantinides, Milton Harris, and René Stulz, eds., *Handbook of the Economics of Finance*, volume vol 2 (North-Holland, Amsterdam).

Diamond, Douglas, and Raghuram Rajan, 2011, Fear of fire sales, illiquidity seeking, and credit freezes, *Quarterly Journal of Economics* 76, 557–591.

Diamond, Douglas W., 1984, Financial intermediation and delegated monitoring, *The Review of Economic Studies* 51, 393–414.

Döskeland, Trond, and Per Strömberg, 2018, Evaluating investments in unlisted equity for the norwegian government pension fund global (gpfg), Technical report, Norwegian Ministry of Finance.

Dyck, Alexander, and Lukasz Pomorski, 2016, Investor scale and performance in private equity investments, *Review of Finance* 20, 1081–1106.

Ewens, Michael, Charles Jones, and Matthew Rhodes-Kropf, 2013, The price of diversifiable risk in venture capital and private equity, *Review of Financial Studies* 26, 1853–1889.

Fitzpatrick, Nick, 2011, Insurance companies: investing their way to solvency.

Giommetti, Nicola, and Morten Sørensen, 2019, Optimal allocation to private equity, Working paper, Copenhagen Business School.

Glode, Vincent, and Richard Green, 2011, Information spillover and performance persistence for hedge funds, *Journal of Financial Economics* 101, 1–17.

Gompers, Paul, and Josh Lerner, 1996, The use of covenants: An analysis of venture partnership agreements, *Journal of Law and Economics* 39, 463–498.

Gompers, Paul, and Josh Lerner, 2000, Money chasing deals? the impact of fund inflows on private equity valuations, *Journal of Financial Economics* 55, 281–325.

Haddad, Valentin, Erik Loualiche, and Matthew Plosser, 2017, Buyout activity: The impact of aggregate discount rates, *Journal of Finance* 72, 371–414.

Harris, Robert S., Tim Jenkinson, and Steven N. Kaplan, 2014a, Private equity performance: What do we know?, *The Journal of Finance* 69, 1851–1882.

Harris, Robert S., Tim Jenkinson, Steven N. Kaplan, and Rüdiger Stucke, 2014b, Has persistence persisted in private equity? evidence from buyout and venture capital funds, Working paper.

Hochberg, Yael V., Alexander Ljungqvist, and Annette Vissing-Jörgensen, 2014, Informational holdup and performance persistence in venture capital, *The Review of Financial Studies* 27, 102–152.

Holmström, Bengt, and Jean Tirole, 1997, Financial intermediation, loanable funds, and the real sector, *The Quarterly Journal of Economics* 112, 663–691.

Innes, Robert D., 1990, Limited liability and incentive contracting with ex-ante action choices, *Journal of Economic Theory* 52, 45–67.

Jensen, Michael, 1989, The eclipse of the public corporation, *Harvard Business Review* 67, 61–74.

Kaplan, Steven, and Jeremy Stein, 1993, The evolution of buyout pricing and financial structure in the late 1980s, *Quarterly Journal of Economics* 108, 313–357.

Kaplan, Steven N., and Antoinette Schoar, 2005, Private equity performance: Returns, persistence, and capital flows, *The Journal of Finance* 60, 1791–1823.

Kaplan, Steven N., and Per Strömberg, 2009, Leveraged buyouts and private equity, *Journal of Economic Perspectives* 23, 121–146.

Laux, Christian, 2001, Limited-liability and incentive contracting with multiple projects, *Rand Journal of Economics* 32, 514–526.

Lerner, Josh, and Antoinette Schoar, 2004, The illiquidity puzzle: theory and evidence from private equity, *Journal of Financial Economics* 72, 3 – 40.

Lerner, Josh, Antoinette Schoar, and Wan Wongsunwai, 2007, Smart institutions, foolish choices: The limited partner performance puzzle, *Journal of Finance* 62, 731–764.

Litvak, K., 2004, Governing by exit: Default penalties and walkaway options in venture capital partnership agreements, *Willamette Law Review* 40, 771–812.

Ljungqvist, Alexander, Matthew Richardson, and Daniel Wolfenzon, 2017, The investment behavior of buyout funds: Theory and evidence, *Financial Management* forthcoming.

Longstaff, Francis, 2004, The flight-to-liquidity premium in u.s. treasury bond prices, *Journal of Business* 77.

Metrick, Andrew, and Ayako Yasuda, 2010, The economics of private equity funds, *Review of Financial Studies* 23, 2303–2341.

Nadauld, Taylor, Berk Sensoy, Keith Vorkink, and Michael Weisbach, 2018, The liquidity cost of private equity investments: Evidence from secondary market transactions, *Journal of Financial Economics* forthcoming.

Phalippou, Ludovic, and Mark M. Westerfield, 2014, Capital commitment and illiquidity risks in private equity, Working paper, University of Oxford Said Business School.

Robinson, David, and Berk Sensoy, 2013, Do private equity managers earn their fees? compensation, ownership, and cash flow performance, *Review of Financial Studies* 26, 2760–2797.

Robinson, David, and Berk Sensoy, 2016, Cyclicality, performance measurement, and cash flow liquidity in private equity, *Journal of Financial Economics* 122, 521–543.

Sahlman, William, 1990, The structure and governance of venture-capital organizations, *Journal of Financial Economics* 27, 473–521.

Sensoy, Berk, Yingdi Wang, and Michael Weisbach, 2014, Limited partner performance and the maturing of the private equity industry, *Journal of Financial Economics* 112, 320–343.

Sørensen, Morten, Neng Wang, and Jinqiang Yang, 2014, Valuing private equity, *The Review of Financial Studies* 27, 1977–2021.