Theoretical Studies of Tournaments and the Venture Capital Industry

Fang, Dawei

Keble College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy in Economics

Trinity Term 2013

Theoretical Studies of Tournaments and the Venture Capital Industry

Fang, Dawei
Keble College, University of Oxford
Thesis submitted for the degree of Doctor of Philosophy in Economics
Trinity Term 2013

Abstract

This thesis consists of three essays. In essay 1, we examine the optimal contract design in the venture capital (VC) industry when a general partner (GP) cannot commit to putting effort into an existing partnership. We show that the first-best contract, a layered debt issued to investors, induces the GP's efficient investment decision and prevents him from diverting effort from the partnership, but its feasibility depends on economic conditions and on whether the GP's ability is observed. The model suggests a procyclical pattern of investment funding and a countercyclical pattern of industrial performance. Moreover, new GPs' fundraising and performance are more sensitive to business cycles than those of established ones.

In essay 2, we study investment duration of a VC fund based on a simple agency conflict between a GP and investors. A short duration financing arrangement can minimize agency conflicts but may offer investors rents. When the rents offered to investors are too large, which may occur when the rate of return of a successful investment is high, the GP has an incentive to use less efficient but rent-saving financing arrangements. We show that there can be cases in which the GP uses a long investment duration financing arrangement or is willing to have the fund capital-constrained.

In essay 3, we study contests where, subject only to a capacity constraint on mean performance, contestants, facing a rank-dependent payoff function, choose arbitrary performance distributions. In the case of symmetric capacity, we derive closed-form solutions for equilibrium performance distributions and analyze the effect of contest structure on equilibrium behavior. We show that equilibrium performance distributions are never dispersion-maximizing and are always right-skewed when the contest is selective. When contestants' capacities are private information, contests serve as a selection mechanism. We analyze the effect of changing contest parameters on strategies, payoffs, and contest selection efficiency and study the selection properties of different contest designs.

Contents

Acknowledgments

I would like to express my gratitude to my thesis supervisors, Professor John Quah (core supervisor) and Professor Thomas Noe (joint supervisor). It would not have been possible for me to complete this thesis without their kind assistance, invaluable suggestions, continuous support, and excellent guidance and direction in every possible way.

When writing this thesis, I have also benefited from the help of various people. I am grateful to Dr. Margaret Meyer for the useful comments and the support and encouragement she gave me in the very stressful period of my life. I would like to thank Einar Bakke, Denis Gromb, Alexander Guembel, Chris Jones, Rafael de Matta, Ines Moreno de Barreda, and Allan Morrison for helpful comments and suggestions.

Last but not least, I would like to thank my friends, my parents, Yongchen Gao and Shinan Fang, and my wife, Kathy for their constant moral support.

Word Count

The official word count is 51,208 words, using number of words, 346, in a 'typical' page (page 3) multiplied by 148 pages.

Essay 1

A Theory of Compensation, Performance, and Fundraising in the Venture Capital Industry

1.1 Introduction

1.1.1 Motivation

Venture capital (VC) investing is typically carried out through a limited partnership structure in which a venture capitalist serves as a general partner (GP). The GP sets up VC funds by raising external capital from investors, who are limited partners (LPs). The GP then invests the funds in projects. By using this limited partnership structure, LPs are not involved in the day-to-day operation of the fund; the GP takes on the administration of the fund with full liability. The GP can approach hundreds of projects every year and it is the GP's decision which projects to invest in, how much to invest, and whether a project should be liquidated or continued. To induce the GP's efficient investment decisions, performance-based GP compensation is typically used to help align the two parties' interests.

In practice, performance-based GP compensation is specified by a nonlinear sharing rule. Typically, the GP earns no cash flows until LPs get back the *nominal* value of the committed fund capital plus a "preferred return"; then the GP receives all the marginal cash flows within a small cash flow region called "GP's *Carried Interest Catch Up*" region; finally the GP and LPs split the marginal cash flows, usually 20-80, above that region (see Sahlman 1990; Gompers and Lerner 1999a). However, the actual compensation structure is not uniform across funds and is more complicated, which includes management fees, timing, and other rules governing distribution of the GP's share of a fund's profits (see Litvak 2009; Phalippou 2009). Nevertheless, it is common that the sharing rule has both convex and concave regions and has a *debt feature*: the GP obtains limited or no performance-based payment before LPs get back their committed fund capital in *net present value* (NPV) terms.[1]

It is puzzling that the two parties do not use equity to align their interests, but use a contract that contains a debt component, which makes the GP bear little downside risk and may thus induce the GP to invest in risky but inefficient projects. To the best of our knowledge, the only existing literature that offers an explanation for this puzzle is Axelson, Strömberg, and Weisbach (2009), who explain it by adverse selection. In their model, there are an infinite supply of fly-by-night operators who cannot bring success to any project they undertake but can store money at the riskless interest rate. Investors cannot distinguish fly-by-night operators from capable venture capitalists. In order to separate from fly-by-night operators, capable venture capitalists use a debt-like contract that offers fly-by-night operators no profits from storing money. However, such an explanation crucially depends on the assumption that without contracts as a signalling device, investors cannot identify real venture capitalists. This assumption may fit new

[1]If LPs' "preferred return" is less than the riskless interest rate, the GP earns positive performance-based compensation even if LPs lose money in NPV terms.

GPs who have no track records of investments, whereas for GPs who have established track records of investments, it is unlikely that investors cannot distinguish them from fly-by-night operators.

In this paper, we explain this debt feature of established GPs' compensation by studying the GP's *effort diversion* problem: the GP cannot commit to devoting his effort to the existing partnership's project; indeed, he could potentially create yet another partnership and devote effort to the new partnership/project, thus ruining the prospect of the existing partnership.

Gompers and Lerner (1996) document that as a GP's effort is crucial for the success of a project, such an effort diversion problem is an important concern of LPs, and to mitigate this problem, a GP and LPs may use covenants included in partnership agreements that restrict the GP's fundraising activities. If the GP is prohibited from raising additional funds, the GP lack money to invest in outside projects and thus will not divert effort into them.

However, it is puzzling that a significant proportion of VC funds do not use such covenants at all. For example, 42% of 140 VC funds in the sample collected by Gompers and Lerner (1996) and 38% of 50 VC funds in the sample collected by Cumming and Johan (2009) do not contain covenants with fundraising restrictions. Several factors may contribute to this phenomenon. For example, covenants cannot be effective without LPs monitoring the GP, which is costly to LPs, and since there are typically multiple LPs in a partnership, they often face the free-rider problem that inhibits effective monitoring. Moreover, covenants may not be effective without a good legal environment, which is suggested by the findings of Cumming and Johan (2009) that the poorer the quality of a country's law, the less frequent the use of covenants in VC funds. Further, the use of covenants may create other agency problems. For instance, covenants that restrict GP's fundraising are typically set in a way that prohibits GP's fundraising until a set percentage, e.g., 70%, of the committed fund capital has been invested. In order to meet this requirement for fundraising, the GP may rush the investment process of the existing fund without putting decent effort into the investment. All these factors may limit the effectiveness of covenants.

In this paper, we show that the effort diversion problem can be resolved through *financial contract design*, and the commonly observed debt feature of GP compensation contributes to the resolution of this problem .

1.1.2 A Brief Discussion of the Model

We consider a penniless GP who has an infinite number of homogeneous project opportunities but only one unit of indivisible effort. We assume that the GP's effort is crucial for the success of a project; all invested projects without the GP's effort will result in negative net profit for sure. If the GP can commit to establishing only one partnership to fund one project, to induce the GP's effort, the partnership's contract just needs to guarantee that the GP's *effort benefit*, which is his money earning difference between working and shirking in the partnership, exceeds his effort cost. However, when the GP cannot commit to establishing just one partnership, this condition is no longer sufficient to induce the GP's effort. When making effort decisions, the GP not only compares his effort benefit with his effort cost, but more importantly, he compares his effort benefit with his *opportunity cost*, which is his effort benefit from applying effort to an outside project by forming another partnership. Thus, to induce the GP's effort, the existing partnership must use an *effort-benefit-maximizing (EBMax) contract* which offers the GP the highest possible effort benefit (subject to investors' break-even condition), since otherwise the GP's effort will be tempted away later on by another partnership which offers the GP a higher effort benefit. By using an EBMax contract, the existing partnership prevents any potential future partnership from (strictly) winning the GP's effort and resolves effort diversion by deterring outside investors from forming other partnerships with the GP.

The derivation of the EBMax contract requires an investigation of the second agency problem, which is *project-choice distortion*. We assume that, with positive probability, a project will have negative NPV even after the application of effort. This probability increases when economic conditions deteriorates. Efficiency would require that the GP should *not* go ahead with the negative-NPV project; he should instead store money in a safe asset. However, the GP will undertake the negative-NPV project unless he is sufficiently paid by storing money in a safe asset. Sufficiently rewarding the GP for storing money raises the GP's earnings from working (as it eliminates the efficiency loss from project-choice distortion), but it also raises the GP's earnings from shirking (as the storage technology is available even if the GP shirks). We find that when the probability of a negative-NPV project is high, the first consideration dominates the second and the EBMax contract is a *layered debt* issued to investors which resolves project-choice distortion by offering the GP a positive payoff from storing money. When the probability of a negative-NPV project is low, the second consideration dominates and the EBMax contract becomes a *debt-like contract* which gives the GP zero payoff from storing money and causes project-choice distortion.

Note that in the above discussion, we assume that the GP bears no adverse selection. Thus, the above discussion fits *established* GPs. For *new* GPs, we assume that they are everything like established GPs except that they bear an extra problem – adverse selection, introduced in the same fashion as Axelson, Strömberg, and Weisbach (2009) discussed before. With adverse selection, new GPs can only raise funds with a debt-like contract, which causes project-choice distortion. When economic conditions are sufficiently bad (the probability of a negative-NPV project is sufficiently high), the efficiency loss from project-choice distortion will be so large that investors can never be fully compensated ex ante. In this case, new GPs cannot get funded.

Combining the established GP's case with the new GP's case gives us an overall picture of the impact of business cycles on the VC industry. When economic conditions are good, both types of GPs get funded with a debt-like contract, in which case all

GPs overinvest and have the same expected performance. When economic conditions become worse, established GPs raise funds with a layered debt contract, which induces efficient investment decisions from them, while new GPs can only raise funds with a debt-like contract which causes project-choice distortion and they are thus outperformed by established GPs. When economic conditions become sufficiently bad, established GPs can still raise funds with a layered debt contract, while new GPs cannot get funded.

Based on these results, we make several interesting predictions, most of which receive empirical support from the existing literature. We predict that the performance and fundraising of established GPs are less sensitive to cycles than those of new GPs, which is consistent with the evidence from Kaplan and Schoar (2005). Moreover, as established GPs' risk-shifting tendencies can only be mitigated in bad times whereas new GPs' risk-shifting tendencies can never be mitigated, we predict that new GPs' investment decisions are less sensitive to cycles and are on average riskier compared to those of established ones. This prediction is consistent with the evidence from Ljungqvist, Richardson, and Wolfenzon (2008). In addition, we predict that the capital inflows of the VC industry are procyclical, which is empirically supported by Gompers and Lerner (1999b) and Robinson and Sensoy (2011), whereas the VC investment has some countercyclical pattern, i.e., the average quality of investments taken in bad times exceeds that of those taken in good times, which coincides with the evidence from Gompers and Lerner (2000).

1.1.3 Related Literature

This paper is most closely related to the literature studying the contractual relationship between a venture capitalist and investors. Axelson, Strömberg, and Weisbach (2009) study the problems of project-choice distortion and adverse selection in the VC financing arrangements under the case where a GP undertakes two consecutive projects without the possibility of effort diversion. In their model, investment pooling can mitigate project-choice distortion but the first best is never attainable. In contrast, this paper replaces the adverse selection problem by the effort diversion problem for established GPs and finds that for established GPs, when economic conditions are adverse, the first best is attainable.

Campello and Matta (2010) develop a model showing how investors, a venture capitalist, and an entrepreneur integrate into a VC fund. In their model, the venture capitalist plays a role as a consultant. Investors can either directly finance the entrepreneur or buy the consulting service from the venture capitalist and then finance the entrepreneur. They find that the service provided by the venture capitalist can increase both ex ante and ex post efficiency, and the price of the service, which can be judged as management fees, is procyclical. In contrast, in our model, the effort of a venture capitalist is crucial for the success of a project, and we focus on the performance-based GP compensation rather than the performance-independent management fees.

1.1.4 Organization of the Paper

The remainder of the paper is structured as follows. Section 1.2 describes the model. Section 1.3 analyzes the equilibrium outcomes for established GPs. Section 1.4 examines the equilibrium outcomes for new GPs. Section 1.5 contrasts the results derived in Sections 1.3 and 1.4 and combines them for the study of the impact of business cycles on the VC industry. Section 1.6 concludes.

1.2 The Model

There are four types of players in this model: established GPs, new GPs, investors, and fly-by-night operators. All players are risk-neutral and have access to a storage technology yielding the riskless interest rate, which we assume to be zero.

The only difference between the two types of GPs is that investors can only distinguish fly-by-night operators from established GPs but not from new GPs. The model described below applies to both types of GPs. For ease of exposition, in the rest of this section "a/the GP" stands for every single GP regardless of the type.

We assume that a GP is penniless but has an infinite number of project opportunities while only one unit of indivisible effort. Effort cost is assumed to be zero. All project opportunities are homogeneous ex ante, and each requires the GP's unit effort followed by investment of I raised from investors, where $I > 0$, to have a chance of success.[2] Note that as the GP only has one unit of effort, he could only bring success to one project at most. Admittedly, in practice, a GP manages a portfolio of projects. As discussed in the earlier part of the paper, resolving effort diversion requires the use of a contract that maximizes the GP's effort benefit, which may result in project-choice distortion. To focus on this mechanism, which still exists even if we allow the GP to manage a portfolio of projects, we abstract from reality by assuming that the GP can only successfully handle one project at most. We assume that the capital market is perfectly competitive, so investors are willing to provide capital as long as they can at least break even. We assume that there is an infinite number of fly-by-night operators who can store money at the riskless interest rate, but apart from that, they can only find real investments that have a maximum payoff less than capital invested.

A brief description of the timing of the model is as follows. A GP raises capital at $t = 0$, makes effort decision at $t = 1$, and makes investment decision at $t = 2$. Cash flows are realized and distributed at $t = 3$.

Now we describe each stage in detail. At $t = 0$, to raise I to invest in a project, a GP offers investors a security which specifies the investors' payoff $s(x)$ backed by the final cash flow x from this investment. To simplify the argument, we assume that a GP does not raise more than I for a project/partnership.[3] We also assume that the GP's

[2] This assumption captures the feature of VC investing that in early stages of a project, capital input is low while labor input is high (see Sahlman 1990; Chemmanur, Krishnan, and Nandy 2011).

[3] In a setting of dynamic contracting with renegotiation, in equilibrium the agent may have a financial slack to deal with the cash flow risk (see DeMarzo and Fishman 2007). However, in our model, there are no consecutive investments, so there is no such a problem. If we allow the GP to raise more than I

effort is unobservable and the total number of partnerships is noncontractible, so s is not contingent on effort or the number of partnerships. Besides, s has to satisfy the following conditions:

Limited Liability (LL): $0 \leq s(x) \leq x$, for all $x \geq 0$;

Monotonicity (M): $s(x)$ and $x - s(x)$ are both nondecreasing in x, for all $x \geq 0$.

These assumptions are quite common in the financial literature on security design (see Harris and Raviv 1989; Nachman and Noe 1994). (M) guarantees that the GP cannot increase his own payoff by manipulating the ex post cash flows, either through borrowing money from a third party or burning money.

The set $S = \{s : R_+ \to R : s \text{ satisfies (LL) and (M)}\}$ is the set of admissible security designs.

After observing the GP's offer, investors decide whether to accept the offer to form a partnership. After investors have made their decisions, the GP can always make another offer in an attempt to form another partnership to undertake another project and cannot commit to establishing any particular number of partnerships. Investors observe the latest offer and established partnerships' securities and decide whether to accept the latest offer. Afterward, again the GP can make another offer for another project, and this fundraising process continues until the GP voluntarily stops making offers.[4] The GP can only form partnerships in sequence. Each partnership's project can only be invested by the money of its partnership. The cash flows of partnerships remain separable from each other. For

for a partnership, there is no change in the qualitative nature of the results, but it adds complexity to characterize the equilibrium security.

[4]If the GP faces an exogenous restriction that he could only form k partnerships at most, then the GP can make an effort commitment to the last partnership by forming the earlier $k-1$ partnerships with securities that pledge all cash flows to investors. Thus, to give effort diversion a nontrivial impact on contracting, we place no upper bound on the number of partnerships the GP could potentially form. However, if we assume that investors incur some labor costs when forming a partnership with the GP, then this "no-upper-bound" assumption is not necessary to make effort diversion a nontrivial problem, since costly contracting ensures that investors can only break even if they can capture the GP's effort. As this alternative assumption adds one more variable, investors' labor costs, into the model, we take the "no-upper-bound" assumption instead. There will be no change in the qualitative nature of our results if we adopt the alternative assumption.

the ith partnership, s_i is contingent only on the cash flows from the ith partnership.

At $t = 1$, the GP decides to which partnership's project he will apply his unit effort. We assume that projects without the GP's effort will be money-burning (MB) projects; if the GP invests I in a money-burning project later on, the gross return from the project will be strictly less than I. If a project obtains the GP's effort, it will be good (G) with probability λ and be bad (B) with probability $1 - \lambda$, where λ is known to all players at $t = 0$ and $0 < \lambda < 1$.[5]

At $t = 2$, the type of each partnership's project is observed by the GP but not by investors. For each partnership, the GP decides whether to invest I to continue its project or store I in a safe asset. We assume that there are no gambling opportunities that can possibly result in an ex post cash flow above I. If continued, a project has a nonnegative ex post cash flow $\tilde{X}_T$ drawn from the cumulative distribution function $T(\cdot)$ known to all players at $t = 0$, where T denotes the type of a project and $T \in \{G, B, MB\}$. Both $B(\cdot)$ and $G(\cdot)$ are assumed to be twice continuously differentiable with $0 \leq G(0) \leq B(0)$ and with their supports contained in $[0, R]$, where $R < \infty$.

We assume that a bad project has negative NPV while a good one has positive NPV, i.e., $0 < E_B[x] < I < E_G[x] < \infty$, where E is the expectations operator and the subscript indicates the type of a project. To distinguish bad projects from money-burning projects, we assume that a bad project can possibly end in a cash flow above I, i.e., $Pr\left\{\tilde{X}_B > I\right\} > 0$.

To arrive at a tractable solution to the problem, we assume that $\tilde{X}_G$ and $\tilde{X}_B$ are

[5] Basically, to obtain the same qualitative results, we only require that the effort cost function and the revenue function are both convex in effort and the optimal effort level exists. These assumptions characterize a situation where focusing on one project is more efficient than spreading effort over projects. Convex revenue functions are widely observed in reality. Although the quality of a project is normally concave in effort, the improved quality can be much valued by consumers, which leads to a convex revenue function.

strictly hazard rate ordered, i.e.,

$$\frac{1 - G(x)}{1 - B(x)} \text{ is strictly increasing in } x \text{ whenever } B(x) \neq 1.$$[6]

Roughly speaking, this means that a good project has higher upside cash flow potential than a bad project.

At $t = 3$, the final cash flow of each project/partnership is realized and distributed. Partnerships are then dissolved.

1.3 Equilibrium Outcomes for an Established GP

In our model, there is no interaction between GPs, so GPs can be studied separately from each other. In this section, we analyze the case of an established GP. For the sake of brevity, "a/the GP" refers to an established GP in this section.

In the first-best world, a GP raises enough capital to undertake one project, puts effort into that project, and continues it if and only if it is a good one. If the GP can commit to establishing just one partnership, as we have assumed zero effort cost, the GP can commit to putting effort into that partnership's project. In this case, the first best is attainable by using a contract, such as equity, that gives the GP sufficient earnings from storing money in a safe asset so that the GP has no incentive to undertake a bad project. However, if the one-partnership commitment cannot be made, such earnings may induce the GP's effort-diverting tendency, in which case the first best may be unattainable.

It is helpful to point out here the difference between effort diversion and the moral hazard of underinvestment of effort. The latter refers to an agent's tendency of shirking in the absence of sufficient effort incentives and has been studied extensively by the existing literature, e.g., Mirrlees (1976), Holmström (1979), and Grossman and Hart (1983).

[6] The strict hazard rate ordering is employed extensively in the economics literature (see, for example, Grossman and Hart 1982 and Noe and Rebello 1992).

Both problems are due to the unobservability and thus non-contractibility of effort. But unlike the moral hazard of underinvestment of effort, effort diversion *per se* brings no ex post inefficiency; if the GP diverts effort from partnership A to partnership B, there is no underinvestment of effort *in aggregate*. However, in what follows, by systematically analyzing the equilibrium outcome for an established GP, we show that effort diversion can sometimes create inefficiency *indirectly* through GP's project-choice distortion.

1.3.1 Equilibrium Conditions for the Established GP's Case

We assume that at $t = 0$ when offering a security gives the GP no more profit, he chooses not to offer. We also assume that at $t = 1$ if there exists more than one partnership to which the GP is willing to devote effort, each of these partnerships will capture the GP's effort with positive probability. Finally, we assume that at $t = 2$ when a GP is indifferent between continuing and discontinuing a project, he chooses the action which maximizes the total surplus. We begin to solve for the equilibrium by first analyzing $t = 2$.

The capital investment decision at $t = 2$

Suppose the GP has formed n partnerships after $t = 0$. We call a partnership *the ith partnership* if right before it is established, there are $i - 1$ partnerships in place.

At $t = 2$, for each partnership, the GP decides whether to invest money into its project or store money in a safe asset based on s_i and his observation of T_i. Define a_i as an indicator of the GP's investment decision regarding the ith partnership's project with $a_i = 0$ referring to the GP's action of abandoning the ith project and storing the ith fund in a safe asset and with $a_i = 1$ referring to his action of investing I in the ith project.

For the ith project, the GP chooses the maximum between his payoff from continuing it, which is $E_{T_i}[x - s_i(x)]$, and his payoff from storing money in a safe asset, which is $I - s_i(I)$. Since a money-burning project cannot end in a cash flow above I, according to the monotonicity condition (M), $E_{MB}[x - s(x)] \leq I - s(I)$. Hence $a_i(MB, s_i) = 0$ for all

$s_i \in S$. For good and bad projects, it is obvious that

$$a_i(G, s_i) = \begin{cases} 1 \quad \text{if} \quad E_G[x - s_i(x)] \geq I - s_i(I) \\ 0 \quad \text{if} \quad E_G[x - s_i(x)] < I - s_i(I) \end{cases} \tag{1.1}$$

and

$$a_i(B, s_i) = \begin{cases} 1 \quad \text{if} \quad E_B[x - s_i(x)] > I - s_i(I) \\ 0 \quad \text{if} \quad E_B[x - s_i(x)] \leq I - s_i(I) \end{cases} . \tag{1.2}$$

The effort decision at $t = 1$

At $t = 1$, the GP decides to which partnership he will devote his effort. If he does not devote effort to any partnership, all projects will be money-burning after $t = 1$ and thus at $t = 2$ the GP will deposit all the money in safe assets, in which case the GP's payoff will be

$$\Pi_0(s_1, \cdots, s_n) = \sum_{i=1}^{n} [I - s_i(I)] . \tag{1.3}$$

By putting effort into the ith partnership, the GP not only obtains (1.3) but also obtains some *effort benefit*, which is the difference in the GP's money earning between working and shirking on the ith partnership's project, expressed as

$$\Delta_i(s_i) = \lambda \max\{E_G[x - s_i(x)], I - s_i(I)\} + (1-\lambda) \max\{E_B[x - s_i(x)], I - s_i(I)\} - (I - s_i(I)) . \tag{1.4}$$

With $a_i(G, s_i)$ and $a_i(B, s_i)$ given by (1.1) and (1.2) respectively, Δ_i can be further expressed as

$$\begin{aligned} \Delta_i(s_i) &= \lambda \{E_G [x - s_i(x)]\}^{a_i(G,s_i)} [I - s_i(I)]^{1-a_i(G,s_i)} \\ &+ (1-\lambda) \{E_B [x - s_i(x)]\}^{a_i(B,s_i)} [I - s_i(I)]^{1-a_i(B,s_i)} - (I - s_i(I)) . \end{aligned} \tag{1.5}$$

Obviously, the GP will put effort into a partnership that offers him the largest effort benefit and our first lemma is thus straightforward.

LEMMA 1.1 *Let $\mathcal{N}$ denote the final set of partnerships the GP forms. The ith partnership will capture the GP's effort with positive probability if and only if s_i satisfies*

$$\Delta_i(s_i) = \max\left\{\Delta_j(s_j)\right\}_{j\in\mathcal{N}} \tag{MAX}$$

Lemma 1.1 implies that partnerships are essentially competing in effort benefit provision for the GP's effort. As effort benefit is determined by contracting at $t = 0$, effort diversion has a strong impact on ex ante contracting, which will be analyzed in detail below.

Fundraising at $t = 0$

At $t = 0$, after forming $(i - 1)$ partnerships, where $i \geq 1$, if the GP attempts to form another partnership, he offers investors a security s_i. For s_i to be *an equilibrium security*, it must be offered by the GP and accepted by investors. These two conditions can be satisfied simultaneously only if the GP creates a strictly positive value for the ith partnership, which requires that (i) the GP puts effort into the ith partnership with positive probability and (ii) the GP continues the ith project if the project is good, i.e., $a_i(G, s_i) = 1$. By Lemma 1.1, the former condition is satisfied if and only if s_i satisfies (MAX). By (1.1), the latter condition is satisfied if and only if s_i satisfies the following incentive-compatibility condition:

$$I - s_i(I) \leq E_G[x - s_i(x)] \tag{ICG}$$

As the ith partnership must capture the GP's effort with positive probability in equilibrium, for investors to accept s_i in equilibrium, they must at least break even by capturing the GP's effort. Thus, with $a_i(G, s_i) = 1$ in equilibrium, s_i must satisfy the following investors' break-even condition:

$$\lambda E_G\left[s_i(x)\right]+(1-\lambda)\left[E_B\left[s_i(x)\right]\right]^{a_i(B,s_i)}s_i(I)^{1-a_i(B,s_i)}\geq I, \tag{BE}$$

where $a_i(B, s_i)$ is given by (1.2), indicating the GP's decision on whether to invest in a bad project.

We summarize the above three necessary conditions for s_i to be an equilibrium security in the following lemma.

Lemma 1.2 *For s_i to be part of an equilibrium, it must satisfy conditions (MAX), (ICG), and (BE).*

The necessary conditions stated in Lemma 1.2 will turn out to be sufficient for us to solve for the equilibrium, so we omit writing down the sufficient conditions for an equilibrium security. In the next subsection, we derive the equilibrium by applying these necessary conditions.

1.3.2 Compensation for an Established GP

By Lemma 1.2, in equilibrium, s_i must satisfy (ICG) and (BE) and offer the GP the largest effort benefit compared to the rest of the partnerships, if any, established by the GP. It is thus quite natural for us to pay particular attention to those securities that maximize $\Delta_i(s_i)$ subject to (ICG) and (BE). We call the solution to this program an *EBMax security* for partnership i. In what follows, we first characterize an EBMax security in the next lemma and then based on this characterization, we prove that an equilibrium security must be an EBMax security. Suppressing the index of a partnership's identity does not affect the program, so we drop the subscript i in the characterization of the solution.

Lemma 1.3 *An EBMax security always exists and makes (BE) bind.*

If $(1-\lambda)\left[I-E_B[x]\right] > I - F_1$*, where* F_1 *is given below, there exists a unique EBMax security and this security is a layered debt, denoted by* s^{LD}*, where*

$$s^{LD}(x) = \begin{cases} x & \text{if} \quad 0 \leq x \leq F_1 \\ F_1 & \text{if} \quad F_1 < x < I \\ x + F_1 - I & \text{if} \quad I \leq x \leq F_2 \\ F_1 + F_2 - I & \text{if} \quad F_2 < x \leq R \end{cases} \tag{1.6}$$

with F_1 *and* F_2 *implicitly and uniquely determined by the following:*

$$\begin{cases} \lambda E_G[s^{LD}(x)] + (1-\lambda)s^{LD}(I) = I \\ I - s^{LD}(I) = E_B[x - s^{LD}(x)] \end{cases}. \tag{1.7}$$

If $(1-\lambda)\left[I-E_B[x]\right] < I - F_1$*, there are multiple EBMax securities and they are all debt-like securities. In this case, the set of EBMax securities, denoted by* S_D*, satisfies:*

$$S_D = \{s \in S: \; s(x) = x \; \text{ if } x \leq I, \text{ and } \lambda E_G[s(x)] + (1-\lambda)E_B[s(x)] = I\}. \tag{1.8}$$

If $(1-\lambda)\left[I-E_B[x]\right] = I - F_1$*, the set of EBMax securities is a combination of* S_D *and* s^{LD}*.*

Proof: See Appendix A. □

Lemma 1.3 shows that an EBMax security offers investors no rents even if the GP devotes effort to their partnership, and conditional on parameters, an EBMax security can either be a *layered debt* or a *debt-like security*. A layered debt has the following feature: it assigns all the marginal cash flows in the region $[0, F_1]$ to investors; then the marginal cash flows in (F_1, I) belong to the GP, and the second equation in (1.7) implies that these marginal cash flows are just sufficient to incentivize the GP to abandon a bad project; finally the marginal cash flows in $[I, F_2]$ and those in $(F_2, R]$ are pledged to investors and the GP respectively. If at $t = 1$, the GP devotes effort to a partnership using a layered debt, the GP will make no project-choice distortion at $t = 2$.

Unlike a layered debt, a debt-like security pledges all the marginal cash flows below I to investors. Since the GP obtains nothing from the partnership using a debt-like security by storing its capital in a safe asset, the GP, conditional on putting effort into such a partnership, has incentives to continue its project even if it is bad.

With an EBMax security fully characterized, we would like to show that an equilibrium security is indeed an EBMax security. Our next lemma, which implies that the GP's fundraising activity is restricted by the use of an EBMax security, is helpful for proving that result.

LEMMA 1.4 *After forming a partnership with an EBMax security, the GP stops fundraising.*

Proof: See Appendix A. □

The intuition of Lemma 1.4 is straightforward. A partnership with an EBMax security (weakly) attracts the GP's effort at $t = 1$ even if the GP establishes multiple partnerships. This implies that if the GP seeks to establish more partnerships after forming a partnership by using an EBMax security, he must earn a positive payoff from those additional partnerships without putting any effort into them, which brings loss to investors of those additional partnerships. To be compensated for this loss, investors of those additional partnerships will require a positive net return when the GP puts effort into their partnerships. However, in that case, those additional partnerships will offer the GP an effort benefit smaller than that offered by the established partnership which uses an EBMax security. Thus, those additional partnerships cannot capture the GP's effort and their investors' aforementioned loss will never be compensated for. Thus, investors will not establish more partnerships with the GP after knowing that the GP has already formed a partnership with an EBMax security. Anticipating this, the GP stops fundraising.

Unlike the traditional belief that the GP's fundraising activity is restricted by covenants, what we find out here shows that the GP's fundraising can be well restricted by financial contract design, through the use of an EBMax security that deters outside investors from financing the GP with extra capital.

Lemma 1.4 implies that at the first time the GP offers investors an EBMax security, investors obtain the full knowledge of the GP's final set of partnerships and their associated securities; there will be no future partnerships entering this set. As this latest partnership the GP attempts to establish offers the GP strictly larger effort benefit compared to previous partnerships, if any, investors are willing to form this partnership with the GP because this partnership can guarantee the GP's effort and make investors break even. In this case, any previous partnership cannot possibly capture the GP's effort and thus cannot be formed in equilibrium. Therefore, in equilibrium, it must be that the GP's first offer is an EBMax security. Investors accept this offer and the GP stops fundraising right afterwards. Thus, by Lemma 1.3 which presents an EBMax security, we obtain the main result of this section, which characterizes the equilibrium.

Theorem 1.1 *There exists a unique equilibrium. In this equilibrium, an established GP only forms one partnership, by offering investors s_1, which is an EBMax security. If*

$$(1-\lambda)\left[I - E_B[x]\right] \geq I - F_1, \tag{1.9}$$

where F_1 is determined by (1.6) *and* (1.7)*, $s_1 = s^{LD}$, where s^{LD} is given by (1.6). In this case, there is no project-choice distortion and an established GP captures the first-best surplus, which is*

$$\lambda\left[E_G[x] - I\right]. \tag{1.10}$$

If (1.9) *does not hold, $s_1 \in S_D$, where S_D is given by (1.8). In this case, an established GP continues the project no matter whether it is good or bad and has the expected payoff* *equal to*

$$\lambda E_G[x] + (1 - \lambda)E_B[x] - I. \tag{1.11}$$

Remark: The expected payoff given by (1.11) equals the first-best surplus less the efficiency loss from project-choice distortion. We call it *the second-best surplus.*

Proof: See Appendix A. □

Effort diversion has a strong impact on contracting; it forces the use of an EBMax security. Such an impact is stronger than the impact of the moral hazard of shirking on contracting. If the GP can commit to establishing just one partnership, then to induce the GP's effort, the security only has to guarantee that the GP's effort benefit exceeds his effort cost. As we have assumed zero effort cost, this condition always holds. However, this condition is not sufficient to induce the GP's effort when commitment of no extra fundraising cannot be made, since in the latter case, the GP has an opportunity cost when he devotes effort to the existing partnership's project. Thus, to induce the GP's effort, the existing partnership has to ensure that the GP's effort benefit not only exceeds his effort cost but also exceeds his opportunity cost, represented by his effort benefit from working on another partnership. The competition for the GP's effort between the existing partnership and potential future partnerships leads to the use of an EBMax security by the existing partnership that maximizes the GP's effort benefit, subject to investors' break-even condition.

Note that if we assume a positive effort cost instead of zero effort cost, then to solve for the equilibrium, we need to check whether the effort benefit offered by an EBMax security can cover the GP's effort cost. If the answer is positive, an EBMax security persists in equilibrium; if negative, the GP cannot raise funds. A detailed discussion of this extension is enclosed in Appendix B.

As the GP only establishes one partnership in equilibrium, Theorem 1.1 implies that effort diversion is always resolved in equilibrium. However, the project-choice distortion problem is resolved only when s^{LD} is used, which requires condition (1.9) to be satisfied. Note that the left-hand side of (1.9) is the efficiency loss from project-choice distortion while the right-hand side is the GP's payoff from storing money of the partnership using s^{LD} in a safe asset. Condition (1.9) is satisfied if and only if the efficiency loss from project-choice distortion is greater than the GP's payoff from storing money of the fund using s^{LD}. In the next section, we link the discussion of condition (1.9) to business cycles.

1.3.3 The Impact of Cycles on an Established GP

Recall that λ is the probability of a project having positive-NPV after application of the GP's effort. We think of λ as representing economic conditions, i.e, higher λ, better economic conditions. Such an interpretation is also adopted by Axelson, Strömberg, and Weisbach (2009). The next proposition shows that an established GP makes efficient investment decisions if and only if the economy is sufficiently bad.

PROPOSITION 1.1 *Given $G(\cdot)$, $B(\cdot)$, and I, there exists a cutoff point λ^* such that*

$$(1-\lambda^*)\left[I - E_B[x]\right] = I - F_1(\lambda^*), \tag{1.12}$$

where $F_1(\lambda^*)$ is determined by (1.6) and (1.7) with λ replaced by λ^*.

If $0 < \lambda \leq \lambda^*$, an established GP gets funded by issuing investors s^{LD}, expressed in (1.6), makes efficient capital investment decisions, and captures the first-best surplus given by (1.10).

If $\lambda^* < \lambda < 1$, an established GP gets funded by issuing investors a debt-like security $s \in S_D$, expressed in (1.8), invests in a project no matter whether it is good or bad, and captures the second-best surplus given by (1.11).

Proof: See Appendix A. □

To understand Proposition 1.1 better, notice that to credibly convince investors of no effort diversion, the GP uses an EBMax contract which maximizes the GP's effort benefit, represented by his earning difference between working and shirking in the partnership. Incentivizing the GP to avoid overinvestment by offering him sufficient earnings from storing money in a safe asset positively affects the GP's earnings from working as it eliminates the efficiency loss from overinvestment, but it also positively affects the GP's earnings from shirking. Thus, providing the GP with efficient investment incentives helps maximize effort benefit only if the first effect dominates the second. This requires economic conditions to be sufficiently bad, because only in bad times, the project, after capturing the GP's effort, has a high chance of being a negative-NPV investment, in which case the first effect (improving the GP's earnings from working by eliminating efficiency loss from overinvestment) is most prominent.

Proposition 1.1 predicts that the form of an established GP's compensation is sensitive to business cycles. Particularly, when economic conditions become worse, an established GP is more likely to be able to earn performance-based compensation even when LPs lose money in NPV terms. Recall that in practice, a GP obtains performance-based compensation only if LPs get back their committed fund capital plus some "preferred return" in *nominal* terms. We thus predict that, when economic conditions become sufficiently adverse, the contract between an established GP and LPs may offer LPs a "preferred return" smaller than the riskless interest rate. Testing this prediction calls for future empirical studies.

Note that, by Proposition 1.1, if λ falls from just above λ^* to just below λ^*, the positive impact on an established GP's payoff from resolving the project-choice distortion problem will dominate the negative impact from a small fall in the expected quality of a project. In this case, an established GP is better off by this small downturn of the economy. As a booming economy is often accompanied by a hot market in reality, what we find here may shed some light on the finding of Litvak (2009) that GP compensation is not significantly correlated to the detrended market hotness.

1.4 Equilibrium Outcomes for a New GP

We now turn to the case of a new GP who, unlike an established GP, cannot be distinguished from fly-by-night operators by investors. Since fly-by-night operators can store funds at the riskless interest rate but add no value to any partnership and also since the supply of fly-by-night operators is potentially infinite, if they earn profit, investors must lose money. Thus, any security that offers fly-by-night operators profit by storing money will not be accepted by investors. Therefore, the securities issued by a new GP must satisfy the following no fly-by-night operators condition.

(NF): For invested capital K, $s(x) = x$ whenever $x \leq K$.

By using a security that satisfies (NF), a new GP can guarantee that fly-by-night operators have no incentive to mimic and thus adverse selection is resolved. Within the set of securities that satisfy (NF), $s \in S_D$, expressed in (1.8), offers the GP the largest effort benefit subject to investors' break-even condition (BE), so effort diversion is resolved by using $s \in S_D$, which is a debt-like security.[7] As a debt-like security offers the GP nothing by storing money in a safe asset, it causes project-choice distortion. When economic conditions are sufficiently bad, the efficiency loss from project-choice distortion can be so large that results in a negative total surplus from investment. In this case, a new GP cannot get funded.

We present the equilibrium outcomes for a new GP in the next proposition.

Proposition 1.2 *Given $G(\cdot)$, $B(\cdot)$, and I, there exists a cutoff point $\bar{\lambda}$ defined as*

$$\bar{\lambda} = \frac{I - E_B[x]}{E_G[x] - E_B[x]}. \tag{1.13}$$

[7] For a new GP, effort diversion is not necessary for the use of $s \in S_D$. As long as the new GP has all the bargaining power against investors, maximizing the GP's payoff subject to (NF) and (BE) is sufficient for the use of $s \in S_D$.

If $0 < \lambda \leq \bar{\lambda}$*, a new GP raises no funds.*

If $\bar{\lambda} < \lambda < 1$*, a new GP establishes a partnership with a debt-like contract* $s \in S_D$*, expressed in* (1.8)*, invests in a project no matter whether it is good or bad, and captures the second-best surplus given by* (1.11)*.*

Proof: See Appendix A. $\square$

Unlike established GPs who can raise funds with an efficient contract in bad times, a new GP can never get funded with an efficient contract due to adverse selection. This leads to interesting differences in performance and fundraising patterns between new GPs and established ones, which will be analyzed in the following section.

1.5 The Impact of Business Cycles on the VC Industry

We combine the results derived in Sections 1.3 and 1.4 to examine the impact of business cycles on the VC industry containing both established and new GPs. The two cutoff points, λ^* and $\bar{\lambda}$ defined in Propositions 1.1 and 1.2 respectively, are particularly important. The following lemma gives the order of the two points.

LEMMA 1.5 $\lambda^* > \bar{\lambda}$*, where* λ^* *and* $\bar{\lambda}$ *are defined by* (1.12) *and* (1.13) *respectively.*

Proof: See Appendix A. $\square$

Figure 1.1 depicts the main results from Propositions 1.1 and 1.2. It shows that the two cutoff points, λ^* and $\bar{\lambda}$ divide the interval $(0, 1)$ into three regions. First note that if λ falls from above $\bar{\lambda}$ to below $\bar{\lambda}$, the capital inflows of the VC industry will become lower as new GPs cannot get funded when $\lambda \leq \bar{\lambda}$.

PROPOSITION 1.3 *The capital inflows of the VC industry are lower when* $\lambda \leq \bar{\lambda}$ *than when* $\lambda > \bar{\lambda}$ *as new GPs raise no funds when* $\lambda \leq \bar{\lambda}$*.*

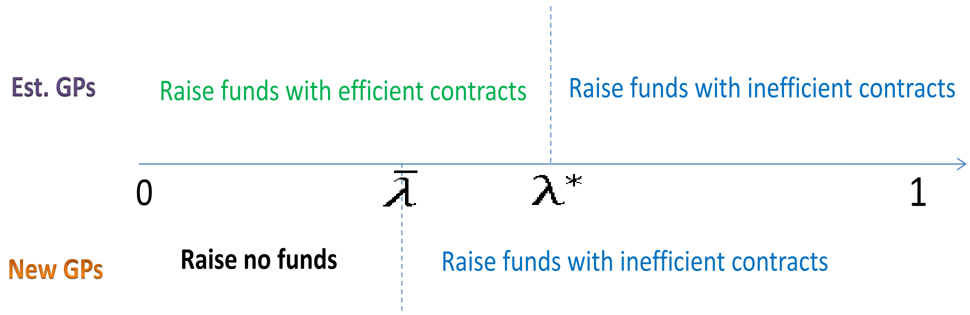


Figure 1.1: New GPs' and Established GPs' Fundraising Patterns

Proof: By Proposition 1.1, established GPs can always get funded. By Proposition 1.2, new GPs get funded if and only if $\lambda > \bar{\lambda}$. $\square$

Proposition 1.3 is consistent with the evidence from Kaplan and Schoar (2005) who show that private equity firms are able to raise funds more easily when overall market conditions are good. Proposition 1.3 also implies that established GPs' fundraising is less sensitive to business cycles than that of new GPs.

Figure 1.1 also implies that established GPs outperform new GPs when $\bar{\lambda} < \lambda \leq \lambda^*$, because when λ falls within this medium range, to raise funds, established GPs use an efficient contract (s^{LD}) while new GPs use an inefficient contract ($s \in S_D$). However, when $\lambda > \lambda^*$, both types raise funds with an inefficient contract ($s \in S_D$), in which case all GPs have the same expected performance. We enclose this result in the following proposition.

Proposition 1.4 *If $\bar{\lambda} < \lambda \leq \lambda^*$, the expected payoff of an established GP is higher than that of a new GP. If $\lambda > \lambda^*$, all GPs have the same expected performance.*

Proof: $\lambda > \lambda^*$ implies $\lambda > \bar{\lambda}$ by Lemma 1.5. The rest of the proof is straightforward by Propositions 1.1 and 1.2, so we omit it. $\square$

Proposition 1.4 implies that the downturn of the economy harms new GPs more than established ones. This implication is consistent with Kaplan and Schoar (2005), who investigate the performance of the private equity partnerships and find that the performance of established GPs is less sensitive to cycles than that of new entrants.

Note also that established GPs only overinvest in good times whereas new GPs always overinvest whenever they can obtain financing. We thus predict that new GPs' investments are less responsive to economic conditions but riskier than those of established ones on average. This prediction receives empirical support from Ljungqvist, Richardson, and Wolfenzon (2008).

As the VC industry exhibits overinvestment when $\lambda > \bar{\lambda}$ but underinvestment when $\lambda \leq \bar{\lambda}$, the next proposition is straightforward.

Proposition 1.5 *The average expected return of invested projects in the industry is higher when* $\lambda \leq \bar{\lambda}$ *than when* $\lambda > \bar{\lambda}$.

Proof: By Propositions 1.1 and 1.2, when $\lambda \leq \bar{\lambda}$, which implies $\lambda < \lambda^*$ by Lemma 1.5, only good projects are invested. When $\lambda > \bar{\lambda}$, new GPs get funded and invest even in bad projects, which reduces the average expected return of invested projects.[8] $\square$

Proposition 1.5 predicts that the average quality of investments taken in bad times exceeds that of those taken in good times. This prediction is broadly consistent with the evidence in Gompers and Lerner (2000).[9] In addition, as we have shown by Proposition 1.3 that VC fundraising is procyclical, Proposition 1.5 further suggests counter-cyclicality in fundraising and performance. Such a pattern is found by Robinson and Sensoy (2011).[10]

[8]When $\lambda > \lambda^*$, established GPs will also overinvest. However, we cannot conclude that the average expected return of invested projects is higher when $\bar{\lambda} < \lambda \leq \lambda^*$ than when $\lambda > \lambda^*$, since although bad projects are more likely to be invested when $\lambda > \lambda^*$, there are less bad projects, which raises the average expected return.

[9]Kaplan and Stein (1993) find evidence suggesting countercyclical investment performance in the buyout market.

[10]Kaplan and Strömberg (2009) find evidence for counter-cyclicality in fundraising and performance in the buyout market.

1.6 Conclusion

It is puzzling that in practice GPs and LPs do not use equity to better align the interests of the two parties but design contracts that have a debt feature, i.e., they offer GPs limited or no performance-based compensation before LPs get back their committed fund capital in NPV terms. It is also puzzling that although LPs are concerned about the effort diversion problem, a significant proportion of partnerships do *not* use covenants that restrict GPs' fundraising activities to mitigate this problem. This paper offers an explanation for both phenomena and shows that the two phenomena are interconnected: the effort diversion problem can be resolved by financial contract design and the commonly observed debt feature of GP compensation can be motivated by the purpose of resolving effort diversion. We also find that resolving effort diversion through financial contract design is not costless; it makes the GP bear less downside risk (compared to equity) and brings in efficiency loss through project-choice distortion when economic conditions are sufficiently good.

The fit and predictive power of the model can be further improved in several ways. First, we can introduce a stronger ex post operating flexibility of the GP into the model. Currently, after the GP puts effort into a project, the project can only take two different distributions of cash flows, each with a certain probability. If the GP has more ex post discretion over the risk choice, e.g., the GP can raise the variance of a project at the cost of a lower mean, it could promote the use of equity as suggested by Admati and Pfleiderer (1994) who find that equity is the only robust financial contract when the space of distributions of cash flows is extremely large. However, equity may not offer the GP the largest effort benefit, so to resolve effort diversion, the optimal contract may be a mix between equity and debt, which may better fit the typical security used in practice described at the beginning of the paper. We leave this line of extension for future research.

Second, it is more realistic to allow the GP to manage a portfolio of projects instead of only one project for a partnership. This enables the GP and LPs to alleviate the project-choice distortion by tying the GP's payoff to his collective performance over several projects even when adverse selection exists (see Axelson, Strömberg, and Weisbach 2009) and thus benefits new GPs. However whether it helps established GPs obtain the first-best surplus requires further study. By Theorem 1.1, the first-best surplus is attainable to established GPs if and only if the efficiency loss from continuing a bad project is larger than a GP's share of a safe asset under an efficient contract. Managing a portfolio of projects has two effects on this condition. First, it reduces the efficiency loss incurred when an established GP puts effort into a partnership with a debt-like contract. This effect makes it more difficult for an established GP to achieve the first best. Second, it reduces the requirement on the GP's share of the safe asset for inducing efficient investment decisions, since the "inside wealth" created by tying investment decisions together has already reduced the limited liability constraint. This second effect makes it easier for an established GP to reach the first best. However, the total effect is not clear and requires further investigation.

Finally, we assumed that contracts of existing partnerships can be observed by all investors. In practice, contract terms of a partnership are often confidential. If we assume that investors other than LPs can only observe the establishment of a partnership without knowing its contract details, investors will form a conjecture on the securities used by the GP's established partnerships when the GP approaches them for financing. It is not difficult to notice that once an established GP has formed one partnership, the only conjecture that is sustainable in equilibrium on the security used by this partnership is an EBMax security. Thus, for an established GP, the equilibrium outcomes are the same as those presented by Proposition 1. For a new GP, to resolve adverse selection, he can only raise funds with a debt-like contract. Because investors do not lose money when the GP diverts effort from their partnership if their partnership uses a debt-like contract, information on the securities used by other partnerships is irrelevant for investors to decide whether to accept a debt-like contract. Hence making security information confidential has no impact on a new GP's financing.

1.7 Appendix A: Proofs

Outline of proof of Lemma 1.3: An EBMax security maximizes $\Delta_i(s_i)$, the GP's effort benefit, subject to (ICG) and (BE). By (1.1), condition (ICG) ensures that an EBMax security has $a_i(G, s_i) = 1$. By replacing $a_i(G, s_i)$ by 1 in $\Delta_i(s_i)$ given by (1.5) and omitting the script i, we obtain that an EBMax security is a solution to the following program, denoted by (MEB):

$$\max_{s(x)\in S} \quad \lambda E_G[x - s(x)] + (1-\lambda)\left\{E_B\left[x - s(x)\right]\right\}^{a(B,s)}\left[I - s(I)\right]^{1-a(B,s)} - (I - s(I))$$

$$\text{s.t.} \qquad \lambda E_G\left[s(x)\right] + (1-\lambda)\left[E_B\left[s(x)\right]\right]^{a(B,s)} s(I)^{1-a(B,s)} \geq I, \qquad \text{(BE)},$$

$$I - s(I) \leq E_G[x - s(x)] \qquad \text{(ICG)}.$$

Note that both the objective function and the (BE) constraint depend on $a(B, s)$, which is determined by s according to (1.2). To derive solutions to (MEB), we first solve for solutions to (MEB) with $a(B, s) = 1$; the associated result is presented in Lemma 1.6 below. We then solve for solutions to (MEB) with $a(B, s) = 0$; the associated result is developed in Lemmas 1.7 and 1.8 below. After that, we complete the proof of Lemma 1.3 by comparing these solutions. Those that give highest effort benefit will be solutions to (MEB).

Lemma 1.6 *With $a(B, s) = 1$, the program* (MEB) *becomes* (MEB1) *as follows:*

$$\max_{s(x)\in S} \quad \lambda E_G[x - s(x)] + (1-\lambda)E_B[x - s(x)] + s(I) - I$$

$$s.t. \qquad \lambda E_G[s(x)] + (1-\lambda)E_B[s(x)] \geq I \qquad \text{(BE1)}$$

$$I - s(I) \leq E_G[x - s(x)] \qquad \text{(ICG)}$$

$$I - s(I) < E_B[x - s(x)] \qquad \text{(B1)}.$$

The solutions to (MEB1) *exist if and only if*

$$\lambda E_G[x] + (1-\lambda)E_B[x] > I, \tag{1.14}$$

in which case the set of solutions is S_D stated in Lemma 1.3. With $s \in S_D$,

$$\Delta(s) = \lambda E_G[x] + (1-\lambda)E_B[x] - I. \tag{1.15}$$

Remark: With $a(G,s) = a(B,s) = 1$, $\Delta(s)$ takes the form of the objective function in (MEB1) according to (1.5), and (BE) becomes (BE1). $a(G,s) = 1$ is guaranteed by (ICG) according to (1.1). $a(B,s) = 1$ is guaranteed by (B1) according to (1.2).

Proof: If $\lambda E_G[x]+(1-\lambda)E_B[x] < I$, (BE1) can never be satisfied as $s(x) \leq x$ according to (LL), so no solution exists in this case.

If $\lambda E_G[x] + (1-\lambda)E_B[x] = I$, (BE1) requires $s(x) = x$ for all $0 \leq x \leq R$, in which case (B1) will be violated, so no solution exists in this case.

Now consider the case where (1.14) holds. The objective function of (MEB1) can be rewritten as

$$\{\lambda E_G[x] + (1-\lambda)E_B[x] - I\} - \{\lambda E_G[s(x)] + (1-\lambda)E_B[s(x)]\} + s(I). \tag{1.16}$$

As the whole term in the first curly bracket of expression 1.16 is independent of s, if there exists a security s that minimizes the whole term in the second curly bracket of 1.16 at the same time maximizes $s(I)$, then this s must maximize 1.16. As s must satisfy (BE1), the whole term in the second curly bracket of 1.16 reaches its minimum when (BE1) binds. By (LL), the maximum value of $s(I)$ equals I, which, by (M), occurs if and only if $s(x) = x$ for all $0 \leq x \leq I$. Therefore, $s \in S_D$, presented by (1.8), maximizes the objective function of (MEB1) subject to (BE1).

Next we prove that $s \in S_D$ satisfies (ICG) and (B1), so it solves (MEB1). The reason is the following. When $s(I) = I$, (ICG) is satisfied as $x \geq s(x)$, and (B1) is satisfied if $E_B[x - s(x)] > 0$. The last condition must hold, since otherwise $s(x) = x$ for all $0 \leq x \leq R$, but as s makes (BE1) bind, this violates (1.14). $\square$

Next we derive solutions to (MEB) when $a(B, s) = 0$. The next lemma will be useful for this derivation, so we prove it first.

LEMMA 1.7 *Given λ, $G(\cdot)$, and $B(\cdot)$, there exists a unique s^{LD} that takes the form of* (1.6) *and satisfies* (1.7). *F_1 is decreasing in λ and is strictly less than I.*

Proof: Since s^{LD} takes the form (1.6) in Lemma 1.3, by plugging s^{LD} into (1.7) and then by integrating by parts and rearranging, (1.7) becomes

$$\begin{cases} \lambda\left[\int_0^{F_1}(1-G(x))dx + \int_I^{F_2}(1-G(x))dx\right] + (1-\lambda)F_1 = I \\ \int_{F_2}^{R}(1-B(x))dx = \int_{F_1}^{I} B(x)dx \end{cases} . \tag{1.17}$$

Let $K(F_1, F_2) \equiv \int_{F_2}^{R}(1-B(x))dx - \int_{F_1}^{I} B(x)dx$, where $0 \leq F_1 \leq I$ and $I \leq F_2 \leq R$. Note that $K(0, I) = -I + E_B[x] < 0$ and $K(I, I) = \int_I^R(1-B(x))dx > 0$. Given F_2, K is continuous and strictly increasing in F_1 for all $F_1 \in [0, I]$, so there exists a unique $\bar{F}_1$ such that $K(\bar{F}_1, I) = 0$, where $0 < \bar{F}_1 < I$, and $K(F_1, I) \geq 0$ for all $F_1 \in [\bar{F}_1, I]$. Note that $K(I, R) = 0$, so $K(F_1, R) \leq 0$ for all $F_1 \in [\bar{F}_1, I]$. Since given F_1, K is continuous and strictly decreasing in F_2, for $F_1 \in [\bar{F}_1, I]$, there must exist a unique F_2, determined by F_1 and denoted by $F_2^K(F_1)$, where $I \leq F_2^K(F_1) \leq R$, such that $K(F_1, F_2^K(F_1)) = 0$. Hence $F_2^K(F_1)$ is well defined when $F_1 \in [\bar{F}_1, I]$. Note that for $F_1 \in [0, \bar{F}_1)$, $K(F_1, F_2) < 0$, so any solution to (1.17) must have $F_1 \in [\bar{F}_1, I]$.

Note that $F_2^K(F_1)$ is increasing in F_1. This is because K is increasing in F_1 but decreasing in F_2, so if $F_1' > F_1$, by $K(F_1, F_2^K(F_1)) = 0$, $K(F_1', F_2^K(F_1)) > 0$, and as $K(F_1', F_2^K(F_1')) = 0$, $F_2^K(F_1') > F_2^K(F_1)$.

Let $J(F_1, \lambda) \equiv \lambda\left[\int_0^{F_1}(1-G(x))dx + \int_I^{F_2^K(F_1)}(1-G(x))dx\right] + (1-\lambda)F_1 - I$, where $\bar{F}_1 \leq F_1 \leq I$. Obviously, J is strictly increasing in F_1. As $K(\bar{F}_1, I) = 0$, $F_2^K(\bar{F}_1) = I$.

Thus $J(\bar{F}_1, \lambda) = -\lambda \int_0^{\bar{F}_1} G(x)dx - (I - \bar{F}_1) < 0$ as $\bar{F}_1 < I$. Since $K(I, R) = 0$, $F_2^K(I) = R$. Hence $J(I, \lambda) = \lambda [E_G[x] - I] > 0$ as $E_G[x] > I$. Since J is strictly increasing in F_1, given λ, there exists a unique $F_1^*(\lambda)$, where $\bar{F}_1 < F_1^*(\lambda) < I$, such that $J(F_1^*(\lambda), \lambda) = 0$. Clearly, when $J = 0$, the first equation of (1.17) is satisfied, and when $K = 0$, the second equation of (1.17) is satisfied. Hence given λ, the unique pair $(F_1^*(\lambda), F_2^K(F_1^*(\lambda)))$ solves (1.17).

By $J(F_1^*(\lambda), \lambda) = 0$ and $F_1^*(\lambda) < I$, $\int_0^{F_1^*(\lambda)}(1 - G(x))dx + \int_I^{F_2^K(F_1^*(\lambda))}(1 - G(x))dx - F_1^*(\lambda) = (I - F_1^*(\lambda))/\lambda > 0$. Hence if $\lambda' > \lambda$, $J(F_1^*(\lambda), \lambda') - J(F_1^*(\lambda), \lambda) = (\lambda' - \lambda)\frac{(I - F_1^*(\lambda))}{\lambda} > 0$, so $J(F_1^*(\lambda), \lambda') > 0$. Besides, $J(F_1^*(\lambda'), \lambda') = 0$, so $F_1^*(\lambda') < F_1^*(\lambda)$ as J is increasing in F_1. Thus, $F_1^*(\lambda)$ is decreasing in λ. $\square$

LEMMA 1.8 *With $a(B, s) = 0$, the program* (MEB) *becomes* (MEB0) *as follows:*

$$\begin{aligned}
\max_{s(x) \in S} \quad & \lambda E_G[x - s(x)] + (1 - \lambda)[I - s(I)] + s(I) - I & \\
s.t. \quad & \lambda E_G[s(x)] + (1 - \lambda)s(I) \geq I & \text{(BE0)} \\
& I - s(I) \leq E_G[x - s(x)] & \text{(ICG)} \\
& I - s(I) \geq E_B[x - s(x)] & \text{(B0)}
\end{aligned}$$

There exists a unique solution to (MEB0)*, which is given by s^{LD} in Lemma 1.3, and*

$$\Delta(s^{LD}) = \lambda E_G[x] - \lambda I + F_1 - I, \tag{1.18}$$

where F_1 is determined implicitly by (1.7)*. $\Delta(s^{LD})$ is strictly increasing in λ.*

Remark: With $a(G, s) = 1$ and $a(B, s) = 0$, $\Delta(s)$ takes the form of the objective function in (MEB0) according to (1.5), and (BE) becomes (BE0). According to (1.1) and (1.2), $a(G, s) = 1$ and $a(B, s) = 0$ are guaranteed by (ICG) and (B0) respectively.

Proof: First note that the objective function of (MEB0) is maximized when $E_G[s(x)] -$

$s(I)$ is minimized, so we rewrite (MEB0) as follows:

$$\min_{s(x)\in S} E_G[s(x)] - s(I) \quad \text{s.t.} \quad \text{(BE0), (ICG) \& (B0).} \tag{1.19}$$

As s^{LD} stated in Lemma 1.3 satisfies (1.7), s^{LD} makes (BE0) and (B0) bind, implied by the first and the second equations in (1.7) respectively. When (B0) binds, (ICG) holds for sure, since $E_B[x-s(x)] \le E_G[x-s(x)]$, implied by strict hazard rate ordering. Thus s^{LD} satisfies (BE0), (ICG), and (B0), implying that these three constraints are not mutually exclusive. Hence (1.19) must have a solution.

According to the restrictions on admissible securities, both $s(x)$ and $x - s(x)$ are nondecreasing in x, so $0 \le s(x) - s(y) \le x - y$, if $x-y \ge 0$. It follows that $|s(x)-s(y)| \le |x-y|$. Therefore, the admissible securities are Lipschitz and thus absolutely continuous (see Royden (1968)). There hence exists a function $w(x) \equiv s'(x)$ almost everywhere on the interval $[0, R]$, such that

$$s(x) = \int_0^x w(r)dr. \tag{1.20}$$

Besides, as $s'(x)$ always lies in $[0,1]$, $w(x)$ can be selected so as to always lie in the same interval.

Note that the constraint (ICG) can be rearranged into $E_G[s(x)] - s(I) \le E_G[x] - I$, which only places an upper bound on the objective function of (1.19), so if (1.19) has a solution, the solution is not determined by (ICG); (ICG) only affects the existence of a solution to (1.19). This motivates us to consider the following relaxed problem:

$$\min_{s(x)\in S} E_G[s(x)] - s(I) \quad \text{s.t.} \quad \text{(BE0) \& (B0).} \tag{1.21}$$

The solutions to (1.19) are the solutions to (1.21) that satisfy (ICG).

By forming Lagrangian using duality theory we can express the problem as:

$$\begin{aligned}\mathcal{L}(s,\zeta_1,\zeta_2) = E_G[s(x)] - s(I) \quad &- \quad \zeta_1 \left\{\lambda E_G[s(x)] + (1-\lambda)s(I) - I\right\} \\ &- \quad \zeta_2 \left\{I - s(I) - E_B[x] + E_B[s(x)]\right\}, \qquad (1.22)\end{aligned}$$

where ζ_1 and ζ_2 are nonnegative scalers and $E_G[s(x)]$ and $E_B[s(x)]$ can be expressed as $\int_0^R w(x)(1-G(x))dx$ and $\int_0^R w(x)(1-B(x))dx$ respectively by integrating by parts and using (1.20), while $s(I)$ can be expressed as $\int_0^I w(x)dx$ by (1.20), so the Lagrangian given by (1.22) can be rewritten and rearranged into

$$\mathcal{L}(w,\zeta_1,\zeta_2) = \int_0^I w(x)(1-B(x))\psi(x)dx + \int_I^R w(x)(1-B(x))\phi(x)dx + \zeta_1 I - \zeta_2 I + \zeta_2 E_B[x], \qquad (1.23)$$

where $\psi(x) \equiv (1-\zeta_1\lambda)\frac{1-G(x)}{1-B(x)} + \frac{-1-\zeta_1(1-\lambda)+\zeta_2}{1-B(x)} - \zeta_2$ and $\phi(x) \equiv (1-\zeta_1\lambda)\frac{1-G(x)}{1-B(x)} - \zeta_2$.

To avoid interrupting the flow of argument, in the following part of the proof, we insert several claims and leave the proofs of these claims after the proof of Lemma 1.8.

Let ζ_1^* and ζ_2^* denote the Lagrange multipliers at optimum, and let $\psi^*(x)$ and $\phi^*(x)$ denote $\psi(x)$ and $\phi(x)$ respectively when $\zeta_1 = \zeta_1^*$ and $\zeta_2 = \zeta_2^*$.

Claim 1: *A solution to* (1.21) *that satisfies* (ICG) *must satisfy*

$$1 - \zeta_1^*\lambda > 0. \qquad (1.24)$$

By Claim 1 and the assumption that $(1-G(x))/(1-B(x))$ is strictly increasing in x, it is obvious that $\phi^*(x)$ is strictly increasing in x.

Claim 2: *There must exist a unique d_2, where $I < d_2 < R$, such that $\phi^*(d_2) = 0$.*

As $\phi^*(x)$ is strictly increasing, Claim 2 implies that when $I \le x < d_2$, $\phi^*(x) < 0$, and when $d_2 < x \le R$, $\phi^*(x) > 0$. Thus, to minimize the Lagrangian (1.23), it is optimal to set

$$w(x) = 1 \quad \text{if} \quad I \le x \le d_2, \quad \text{and} \quad w(x) = 0 \quad \text{if} \quad d_2 < x \le R. \qquad (1.25)$$

Claim 3: *A solution to* (1.21) *that satisfies* (ICG) *must satisfy*

$$-1 - \zeta_1^*(1-\lambda) + \zeta_2^* > 0. \tag{1.26}$$

Note that $\psi^*(x) = \phi^*(x) + \frac{-1-\zeta_1^*(1-\lambda)+\zeta_2^*}{1-B(x)}$. As $\phi^*(x)$ is strictly increasing in x, by Claim 3, $\psi^*(x)$ is strictly increasing in x.

Claim 4: *There must exist a unique* d_1*, where* $0 \le d_1 < I$*, such that* $\psi^*(d_1) = 0$.

As $\psi^*(x)$ is strictly increasing, Claim 4 implies that when $0 \le x < d_1$, $\psi^*(x) < 0$, and when $d_1 < x < I$, $\psi^*(x) > 0$. Thus, to minimize the Lagrangian (1.23), it is optimal to set

$$w(x) = 1 \quad \text{if} \quad 0 \le x \le d_1, \quad \text{and} \quad w(x) = 0 \quad \text{if} \quad d_1 < x < I. \tag{1.27}$$

Since any solution to (1.19) solves (1.21) and satisfies (ICG), it must satisfy (1.25) and (1.27), and thus must take the structure of a layered debt given by (1.6) with F_1 and F_2 replaced by d_1 and d_2 respectively.

To identify d_1 and d_2, next we show that any solution to (1.19) must make (BE0) and (B0) bind, which is equivalent to demonstrating that $\zeta_1^* \neq 0$ and $\zeta_2^* \neq 0$.

If $\zeta_1^* = 0$, then by (1.26), $\zeta_2^* > 1$. Therefore, $\psi^*(x) = (\zeta_2^* B(x) - G(x))/(1-B(x))$, which is strictly greater than $(B(x)-G(x))/(1-B(x))$ for all $0 \le x < I$. Strict hazard rate ordering and the initial condition $B(0) \ge G(0)$ imply $B(x) \ge G(x)$ for all $0 \le x \le R$.[11] Hence $\psi^*(x) > 0$ for all $0 \le x < I$, which violates Claim 4, so $\zeta_1^* \neq 0$, and thus (BE0) must bind.

If $\zeta_2^* = 0$, $\phi^*(x) = (1-\zeta_1^*\lambda)(1-G(x))/(1-B(x))$, which is strictly greater than zero for all $I < x < R$ by (1.24), but this violates Claim 2. Therefore, $\zeta_2^* \neq 0$, and thus (B0) must bind.

To recap, we have shown that any solution to (1.19) must take the form of a layered

[11] By strict hazard rate ordering, $(1-G(x))/(1-B(x)) \ge (1-G(0))/(1-B(0))$ for all $0 \le x \le R$. By the initial condition $B(0) \ge G(0)$, $(1-G(0))/(1-B(0)) \ge 1$. Hence $(1-G(x))/(1-B(x)) \ge 1$ for all $0 \le x \le R$, and thus $B(x) \ge G(x)$ for all $0 \le x \le R$.

debt given by (1.6) with F_1 and F_2 replaced by d_1 and d_2 respectively and make (BE0) and (B0) bind. This implies that the solution must solve (1.6) and (1.7). By Lemma 1.7, there is a unique security, denoted by s^{LD}, that solves (1.6) and (1.7), so s^{LD} is the unique solution to (1.19).

To obtain (1.18), substitute s^{LD} for s in the objective function of (MEB0) and use the facts that s^{LD} makes (BE0) bind and has $s^{LD}(I) = F_1$.

Finally, to prove that $\Delta(s^{LD})$ is strictly increasing in λ, let $s^{LD}_{\lambda'}$ and $s^{LD}_{\lambda''}$ denote the solutions to (MEB0) when $\lambda = \lambda'$ and $\lambda = \lambda''$ respectively, where $\lambda' < \lambda''$. Because $a(G, s^{LD}_{\lambda'}) = 1$ and $a(B, s^{LD}_{\lambda'}) = 0$, by (1.5),

$$\Delta(s^{LD}_{\lambda'}, \lambda') = \lambda' \left\{ E_G[x - s^{LD}_{\lambda'}] - (I - s^{LD}_{\lambda'}(I)) \right\}. \tag{1.28}$$

As $s^{LD}_{\lambda'}$ satisfies (ICG), the whole term in the curly bracket above must be weakly positive. Thus, $\Delta(s^{LD}_{\lambda'}, \lambda')$ must be weakly less than $\lambda'' \left\{ E_G[x - s^{LD}_{\lambda'}] - (I - s^{LD}_{\lambda'}(I)) \right\}$ as $\lambda' < \lambda''$. Hence, $\Delta(s^{LD}_{\lambda'}, \lambda') \leq \Delta(s^{LD}_{\lambda'}, \lambda'')$. Note that when $\lambda = \lambda''$, the solution to (MEB0) is $s^{LD}_{\lambda''}$ but not $s^{LD}_{\lambda'}$. This implies that $\Delta(s^{LD}_{\lambda''}, \lambda'') > \Delta(s^{LD}_{\lambda'}, \lambda'')$. Therefore, $\Delta(s^{LD}_{\lambda''}, \lambda'') > \Delta(s^{LD}_{\lambda'}, \lambda')$. □

Below we prove by contradiction that the four claims hold.

Proof of Claim 1: Suppose (1.24) does not hold. Then $\phi^*(x)$ will be less than zero for all $0 \leq x \leq R$, suggesting that (1.23) is minimized by choosing $w(x) = 1$ for all $I \leq x \leq R$. Thus $E_G[x - s(x)] = \int_0^I (x - s(x))dG(x)$. By (M), $x - s(x)$ is nondecreasing, so $\int_0^I (x - s(x))dG(x) \leq \int_0^I (I - s(I))dG(x)$, which is strictly less than $I - s(I)$. Hence, $E_G[x - s(x)] < I - s(I)$, which violates (ICG). □

Proof of Claim 2: Suppose Claim 2 does not hold. Since $\phi^*(x)$ is continuous and we have shown in the proof of Lemma 1.8 that $\phi^*(x)$ is strictly increasing, it must be either $\phi^*(x) < 0$ for all $I \leq x < R$ or $\phi^*(x) > 0$ for all $I < x \leq R$. In the former case, it is optimal to set $w(x) = 1$ for all $I \leq x < R$, but we can see that this violates (ICG) by applying the same argument in the proof of Claim 1. In the latter case, it is optimal to set $w(x) = 0$ for all $I < x \leq R$, but this violates (BE0) as investors obtain no marginal cash flows from the region above I to compensate for their loss if the cash flow ends in less than I. $\square$

Proof of Claim 3: Suppose Claim 3 does not hold. Claim 2 together with strict monotonicity of $\phi^*(x)$ implies that $\phi^*(x) < 0$ for all $0 \leq x \leq I$. Thus, $\psi^*(x) < 0$ for all $0 \leq x \leq I$. Hence, to minimize (1.23), it is optimal to set $w(x) = 1$ for all $0 \leq x \leq I$, but this violates (B0) unless $s(x) = x$ for all $0 \leq x \leq R$, which condition requires (BE0) to be relaxed. However, when (BE0) is relaxed, $\zeta_1^* = 0$, and thus $\phi^*(x) = [(1 - G(x))/(1 - B(x))] - \zeta_2^*$. By Claim 2, $\phi^*(d_2) = 0$, where $I < d_2 < R$, so $\zeta_2^* = (1 - G(d_2))/(1 - B(d_2))$, which must be strictly greater than 1 by strict hazard rate ordering together with $B(0) \geq G(0)$. But when $\zeta_1^* = 0$ and $\zeta_2^* > 1$, Claim 3 still holds, which contradicts what we just assumed. $\square$

Proof of Claim 4: Suppose Claim 4 does not hold. Since $\psi^*(x)$ is continuous and we have shown in the proof of Lemma 1.8 that $\psi^*(x)$ is strictly increasing, it must be either that $\psi^*(x) < 0$ for all $0 \leq x < I$ or that $\psi^*(x) > 0$ for all $0 \leq x \leq I$.

In the former case, it is optimal to set $w(x) = 1$ for all $0 \leq x < I$, but this violates (B0) unless $s(x) = x$ for all $0 \leq x \leq R$, which condition requires (BE0) to be relaxed, and thus requires $\zeta_1^* = 0$. In the proof of Claim 3, we have shown that when (BE0) is relaxed, $\zeta_2^* > 1$. With $\zeta_1^* = 0$ and $\zeta_2^* > 1$, $\psi^*(x) = (\zeta_2^* B(x) - G(x))/(1 - B(x))$, which must be strictly greater than zero. This contradicts the assumption of this case that $\psi^*(x) < 0$ for all $0 \leq x < I$.

In the latter case, it is optimal to set $w(x) = 0$ for all $0 \leq x \leq I$. This gives $s(I) = 0$, implying that (B0) is relaxed at optimum, and thus $\zeta_2^* = 0$. When $\zeta_2^* = 0$, $\phi^*(x) = (1 - \zeta_1^* \lambda)(1 - G(x))/(1 - B(x))$, which must be strictly greater than zero for all $I < x < R$ according to (1.24). However, this violates Claim 2. $\square$

Proof of Lemma 1.3: First, by Lemma 1.6, when (1.14) is violated, (MEB1) has no solution. In such a case, the solution to (MEB) must be the solution to (MEB0), which is s^{LD} according to Lemma 1.8.

Next, by Lemma 1.6, when (1.14) is satisfied, the effort benefit associated with the solutions to (MEB1) is given by (1.15), while by Lemma 1.8, the effort benefit associated with the solution to (MEB0) is given by (1.18). It is thus straightforward that between the two levels of effort benefits, (1.15) is weakly higher when

$$(1-\lambda)[I - E_B[x]] \leq I - F_1, \tag{1.29}$$

where F_1 is determined implicitly by (1.7). When (1.29) holds with equality, the effort benefit given by (1.15) equals that given by (1.18). In fact, (1.29) implies (1.14), which will be proved soon. Therefore, when (1.29) holds with strict inequality, the solutions to (MEB) are the solutions to (MEB1) and are thus $s \in S_D$. When (1.29) holds with equality, the solutions to (MEB) are both the solutions to (MEB1) and the solution to (MEB0), i.e., both $s \in S_D$ and s^{LD}. When (1.29) is violated, s^{LD} is the only solution to (MEB).

Now we show that (1.29) implies (1.14). Since s^{LD} makes (B0) bind, implying that (ICG) is relaxed, we obtain

$$E_G[x] > I - F_1 + E_G[s^{LD}(x)]. \tag{1.30}$$

By (1.29), $(1-\lambda)E_B[x] \geq F_1 - \lambda I$. Thus, $\lambda E_G[x] + (1-\lambda)E_B[x]$ is strictly greater than $(1-\lambda)F_1 + \lambda E_G[s^{LD}(x)]$, which equals I as $s^{LD}(I) = F_1$ and s^{LD} satisfies (1.7). Hence (1.14) is satisfied. $\square$

Proof of Lemma 1.4: Consider the subgame where the $(i-1)$th partnership has been formed by using s_{i-1}, which is an EBMax security. If the GP makes one more offer s_i, by Lemma 1.2, for s_i to be part of an equilibrium, s_i must satisfy (ICG) and (BE) and gives the GP an effort benefit weakly higher than the effort benefit given by the $(i-1)$th partnership, i.e., $\Delta_i(s_i) \geq \Delta_{i-1}(s_{i-1})$. To satisfy these conditions, s_i must also be an EBMax security. As the ith partnership offers the GP the same amount of effort benefit as that offered by the $(i-1)$th partnership, the ith partnership cannot guarantee the GP's effort. As an EBMax security makes (BE) bind, i.e., investors of the ith partnership break even if they capture the GP's effort, with the GP's effort unsecured, investors are willing to form the ith partnership only if they can also break even when the GP puts no effort into their partnership. Thus, investors accept s_i only if $s_i(I) = I$. Since the $(i-1)$th and the ith partnerships offer the GP the same amount of effort benefit, the GP weakly prefers storing money of the ith partnership in a safe asset. However, by doing so, the GP obtains no more profit because $s_i(I) = I$, so the GP has no incentive to establish the ith partnership. $\square$

Proof of Theorem 1.1: Suppose the GP raises n partnerships in equilibrium. By Lemma 1.2, each partnership's security must satisfy (MAX) condition stated in Lemma 1.1. Thus, the n securities must offer the GP the same amount of effort benefit.

We then prove by contradiction that the n securities must all be EBMax securities. Suppose at least one of them differs from an EBMax security. By Lemma 1.2, this security must satisfy (ICG) and (BE) and thus offers the GP a smaller amount of effort benefit compared to that offered by an EBMax security. As the n securities must offer the GP the same amount of effort benefit, they must all lose the GP's effort when competing against a partnership using an EBMax security. This, together with Lemma 1.4, implies that if the GP offers investors one more security in the form of an EBMax security in an attempt to form the $(n+1)$th partnership, investors know that this $(n+1)$th partnership, if established, can guarantee the GP's effort. As an EBMax security satisfies (ICG) and (BE), investors can break even when the GP's effort is secured, so investors are willing to establish this $(n+1)$th partnership with the GP by using an EBMax security. As the GP strictly prefers putting effort into this partnership rather than into any of the previous

n partnerships, the GP must be made strictly better off by establishing this one more partnership and hence he will deviate by continuing fundraising after the establishment of the n partnerships. We thus obtain a contradiction.

As the n partnerships must all use EBMax securities, by Lemma 1.4, the GP stops fundraising after the establishment of the first partnership. Therefore, $n = 1$ in equilibrium.

Since s_1 is an EBMax security, by Lemma 1.3, when (1.9) holds with strict inequality, $s_1 = s^{LD}$, offering the GP the surplus from continuing only the good project, expressed by (1.10). When (1.9) is violated, $s_1 \in S_D$, offering the GP the surplus from continuing a project for sure, expressed by (1.11). When (1.9) holds with equality, both s^{LD} and $s \in S_D$ are EBMax securities, but the GP chooses s^{LD} as the payoff given by (1.10) is strictly greater than that given by (1.11). $\square$

Proof of Proposition 1.1: Define $Z(\lambda) \equiv (1-\lambda)[I - E_B[x]] - (I - F_1(\lambda))$, where $F_1(\lambda)$ is determined by (1.17). By Lemma 1.7, $F_1(\lambda)$ is decreasing in λ, so obviously Z is strictly decreasing in λ as $I > E_B[x]$.

By the first equation of (1.17), $\lim_{\lambda \to 0} F_1(\lambda) = I$, so $\lim_{\lambda \to 0} Z(\lambda) = \lim_{\lambda \to 0} F_1(\lambda) - E_B[x] = I - E_B[x] > 0$. By Lemma 1.7, $F_1(\lambda) < I$, so $\lim_{\lambda \to 1} Z(\lambda) = \lim_{\lambda \to 1} F_1(\lambda) - I < 0$. As Z is strictly decreasing in λ, there must exist a unique λ^*, where $0 < \lambda^* < 1$, such that $Z(\lambda^*) = 0$. $Z(\lambda) \geq 0$ if and only if $\lambda \leq \lambda^*$. In other words, (1.9) is satisfied if and only if $\lambda \leq \lambda^*$. The rest of the proof follows directly from Theorem 1.1. $\square$

Proof of Proposition 1.2: By (NF), $s(I) = I$, so a new GP earns nothing from storing money and thus will continue a project no matter whether it is good or bad, which results in a total surplus equal to (1.11). If $\lambda \leq \bar{\lambda}$, where $\bar{\lambda}$ is given by (1.13), (1.11) is nonpositive, so no security can make investors break even and a new GP simultaneously earn profit. Thus in such a case, no fund will be raised.

If $\lambda > \bar{\lambda}$, (1.11) is strictly positive. To maximize profit, among all the admissible securities that have $s(I) = I$, a new GP chooses $s \in S_D$ as it makes investors just break even. As $s \in S_D$ causes project-choice distortion, the new GP captures the second-best surplus. Investors break even regardless of the GP's effort, so they accept the offer. Adding more partnerships with $s(I) = I$ gives the GP no more profit, so he stops fundraising after forming the first partnership. $\square$

Proof of Lemma 1.5: Define $\varphi(\lambda) \equiv \lambda E_G[x] + (1-\lambda)E_B[x]$. Define $s^{LD}_{\lambda^*}$ as the s^{LD} contract when $\lambda = \lambda^*$. By (1.12), $\varphi(\lambda^*)$ equals $\lambda^* E_G[x] + F_1(\lambda^*) - \lambda^* I$ and is thus strictly greater than $\lambda^* E_G[s^{LD}_{\lambda^*}(x)] + (1-\lambda^*)F_1(\lambda^*)$ by (1.30). Note that the last expression equals I by the first equation in (1.7) together with $s^{LD}_{\lambda^*}(I) = F_1(\lambda^*)$. Therefore, $\varphi(\lambda^*) > I$. By (1.13), $\varphi(\bar{\lambda}) = I$. Obviously, $\varphi(\lambda)$ is strictly increasing in λ, so $\lambda^* > \bar{\lambda}$. $\square$

1.8 Appendix B: Nonzero Effort Cost

In the main text of the paper, we assumed zero effort cost. Here we relax this assumption by assuming a positive effort cost, $c > 0$, regarding the GP's effort. We assume that the effort cost is not too large, i.e., $c < E_G[x] - I$, since otherwise it is obvious that projects are not economically viable. For simplicity, we suppose that there is no cost variation between a new GP and an established GP. All the rest of the model configurations remain the same as those presented in Sections 1.2 and 1.4.

When $c > 0$, to induce the established GP's effort, the security s not only has to maximize the established GP's effort benefit, $\Delta(s)$, but also has to guarantee that the effort benefit exceeds the effort cost, i.e., $\Delta(s) \geq c$. Thus, we only need to check whether an EBMax security satisfies the latter condition. If the answer is positive, the equilibrium security is still that presented by Theorem 1.1; if otherwise, an established GP cannot raise funds.

For new GPs, adverse selection can only be resolved by using a debt-like security, $s \in S_D$. Thus with $c > 0$, an debt-like security is still the equilibrium security if $\Delta(s) \geq c$, where $s \in S_D$; if otherwise, new GPs cannot raise funds.

The next proposition characterizes the impact of economic conditions on the financing features of both new and established GPs with $c > 0$.

PROPOSITION 1.6 *A new GP gets funded with $s \in S_D$ if and only if*

$$\lambda \geq \frac{I + c - E_B[x]}{E_G[x] - E_B[x]} \equiv \bar{\lambda}_c; \tag{1.31}$$

if otherwise, a new GP cannot raise funds.

Let $\hat{c}$ defined as

$$\hat{c} \equiv [E_G[x] - E_B[x]]\, \lambda^* + E_B[x] - I, \tag{1.32}$$

where λ^* is defined by (1.12).

If $0 < c \leq \hat{c}$, then $\bar{\lambda}_c \leq \lambda^*$. In this case, there exists a cutoff point $\tilde{\lambda}_c$ determined implicitly by

$$\tilde{\lambda}_c E_G[x] - \tilde{\lambda}_c I + F_1(\tilde{\lambda}_c) - I = c, \tag{1.33}$$

where $F_1(\tilde{\lambda}_c)$ is determined by (1.6) and (1.7) with λ replaced by $\tilde{\lambda}_c$. If $\lambda < \tilde{\lambda}_c$, an established GP cannot raise funds. If $\tilde{\lambda}_c \leq \lambda \leq \lambda^*$, an established GP raises funds with s^{LD}. If $\lambda > \lambda^*$, an established GP gets funded with $s \in S_D$.

If $\hat{c} < c < E_G[x] - I$, then $\bar{\lambda}_c > \lambda^*$. In this case, an established GP gets funded with $s \in S_D$ if and only if condition (1.31) is satisfied; if otherwise, an established GP cannot raise funds.

Proof: A new GP raises funds with $s \in S_D$ if and only if

$$\Delta(s) \geq c \qquad \forall \quad s \in S_D. \tag{1.34}$$

When $s \in S_D$, $\Delta(s)$ is given by (1.15). Thus, condition (1.34) is equivalent to condition (1.31). If (1.34) is violated, a new GP will not exert effort, in which case he cannot raise funds.

An established GP raises funds with an EBMax security if and only if an EBMax security results in an effort benefit greater than c. Note that the effort benefit provided by an EBMax security must be the maximum between the effort benefit provided by $s \in S_D$, presented by (1.15), and that by s^{LD}, presented by (1.18). Thus an established GP raises funds with an EBMax security if and only if the following condition is satisfied:

$$\max\{\lambda E_G[x] + (1-\lambda)E_B[x] - I,\ \lambda E_G[x] - \lambda I + F_1(\lambda) - I\} \geq c, \tag{1.35}$$

where $F_1(\lambda)$ is determined implicitly by (1.7).

Note that the first element in the curly bracket is the effort benefit provided by $s \in S_D$, which is strictly increasing in λ as $E_G[x] > E_B[x]$. The second element in the curly bracket is the effort benefit provided by s^{LD}, which has been proved to be strictly increasing in λ (see Lemma 1.8 in Appendix A). Thus, the left-hand side of (1.35) is strictly increasing in λ.

When $c > \hat{c}$, by the definitions of $\hat{c}$ and $\bar{\lambda}_c$ presented by (1.32) and (1.31) respectively, it is obvious that $\bar{\lambda}_c > \lambda^*$. In this case, when $\lambda = \bar{\lambda}_c$, $\lambda > \lambda^*$. Thus, when $\lambda = \bar{\lambda}_c$, $s \in S_D$ offers larger effort benefit compared to s^{LD}. Hence when $\lambda = \bar{\lambda}_c$, the left-hand side of (1.35) equals $\bar{\lambda}_c E_G[x] + (1-\bar{\lambda}_c)E_B[x] - I$. By the definition of $\bar{\lambda}_c$ given by (1.31), condition (1.35) holds at equality when $\lambda = \bar{\lambda}_c$. This implies that, as the left-hand side of (1.35) is strictly increasing in λ, when $\lambda \geq \bar{\lambda}_c$, condition (1.35) is satisfied, in which case an established GP gets funded with $s \in S_D$, while when $\lambda < \bar{\lambda}_c$, condition (1.35) is violated, in which case an established GP cannot get funded.

When $c \leq \hat{c}$, it is obvious that $\bar{\lambda}_c \leq \lambda^*$. In this case, when $\lambda = \bar{\lambda}_c$, the left-hand side of (1.35) must be greater than $\bar{\lambda}_c E_G[x] + (1-\bar{\lambda}_c)E_B[x] - I$ and thus greater than c. As $E_B[x] < I$ and $F_1(\lambda) \leq I$, when $\lambda = 0$, the left-hand side of (1.35) is nonpositive and thus must be strictly less than c, as we assumed $c > 0$. As the left-hand side of (1.35) is strictly increasing in λ, there must exist a unique $\tilde{\lambda}_c$, defined by (1.33), such that condition (1.35) is violated if and only if $\lambda < \tilde{\lambda}_c$, in which case an established GP cannot get funded. If $\lambda \geq \tilde{\lambda}_c$, an established GP's fundraising remains the same as that stated by Proposition 1.1. $\square$

Unsurprisingly, the cutoff point $\bar{\lambda}$, presented by (1.13), coincides with $\bar{\lambda}_c$, presented by (1.31), when $c = 0$. The effort cost $c > 0$ pushes up the threshold in economic conditions for a new GP to obtain financing; the higher the effort cost, c, the higher the threshold, $\bar{\lambda}_c$.

If effort cost is not too large, i.e., if $0 < c \leq \hat{c}$, the only change in an established GP's financing, compared to the model with $c = 0$, is that an established GP cannot raise funds when economic conditions are sufficiently bad, i.e., when $\lambda < \tilde{\lambda}_c$. This case can be further divided into two subcases: when

$$\lambda < \frac{c}{E_G[x] - I} \equiv \lambda_c^o, \tag{1.36}$$

it is economically inefficient to apply effort to any project; when $\lambda_c^o \leq \lambda < \tilde{\lambda}_c$, although a project is economically viable, the effort benefit provided by an EBMax security is not large enough to cover the GP's effort cost, in which case an established GP will shirk after financing and thus investors will not finance him. When economic conditions are sufficiently good, i.e, when $\lambda \geq \tilde{\lambda}_c$, the effort benefit provided by an EBMax security exceeds the GP's effort cost, in which case by using an EBMax security, an established GP will exert effort and, as an EBMax security resolves effort diversion, he will not divert effort. We summarize these observations for the case where $0 < c \leq \hat{c}$ in Figure 1.2.

Figure 1.2 shows that with a positive but not-so-large effort cost, i.e., $0 < c \leq \hat{c}$, established GPs cannot raise funds when economic conditions are too bad, which situation is absent from Figure 1.1. Apart from this difference, the fundraising and performance of new GPs, established GPs, and the overall VC industry feature the same patterns with respect to business cycles as those discussed in Section 1.5.

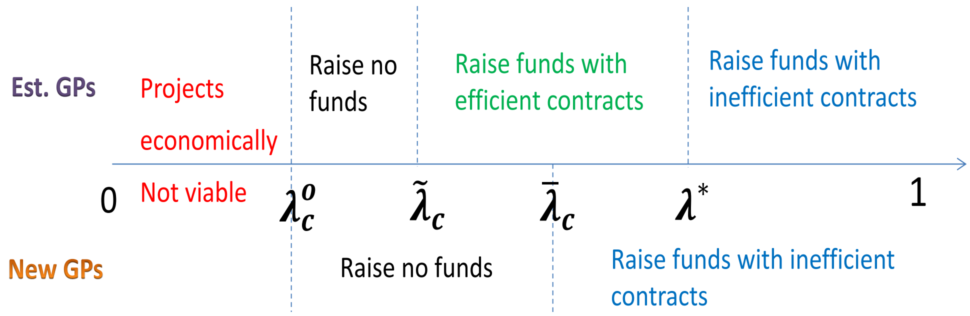


Figure 1.2: New GPs' and established GPs' fundraising patterns when $0 < c \leq \hat{c}$

If effort cost is sufficiently large, i.e., if $c > \hat{c}$, to cover the GP's effort cost, the effort benefit also needs to be sufficiently large, which can only happen when economic conditions are sufficiently good. However, when the economy is good enough, an EBMax security is a debt-like security that induces project-choice distortion. In this case, an established GP's fundraising pattern will be the same as that of a new GP; whenever an established GP obtains financing, he overinvests. Figure 1.3 depicts this case.

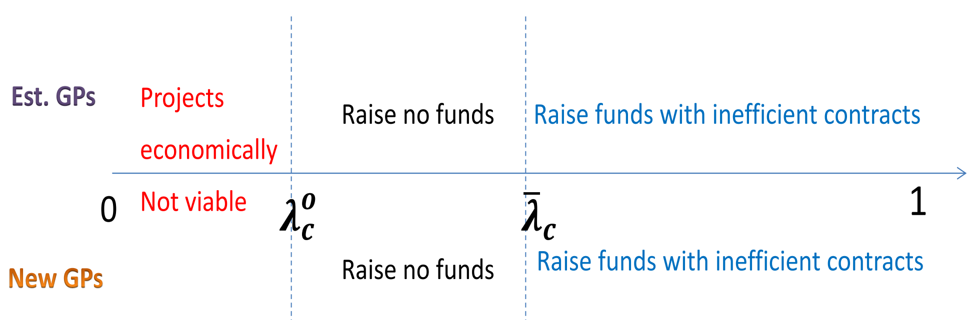


Figure 1.3: New GPs' and established GPs' fundraising patterns when $\hat{c} < c < E_G[x] - I$

Bibliography

[1] Admati, Anat R., and Paul Pfleiderer, 1994, Robust financial contracting and the role of venture capitalists, *Journal of Finance*, 49, 371-402.

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

[3] Campello, Murillo, and Rafael A. da Matta, 2010, How are venture capitalists rewarded? The economics of venture capital partnerships, Working paper.

[4] Chemmanur, Thomas J., Karthik Krishnan, and Debarshi Nandy, 2011, How does venture capital financing improve efficiency in private firms? A look beneath the surface, *Review of Financial Studies* 12, 4037-4090.

[5] Cumming, Douglas J., and Sofia A. Johan, 2009, *Venture Capital and Private Equity Contracting: An International Perspective* (Elsevier Science Adademic Press).

[6] DeMarzo, Peter M., and Michael J. Fishman, 2007, Optimal long-term financial contracting, *Review of Financial Studies* 5, 2079-2128.

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

[8] Gompers, Paul, and Josh Lerner, 1999a, An analysis of compensation in the U.S. venture capital partnership, *Journal of Financial Economics* 51, 3-44.

[9] Gompers, Paul, and Josh Lerner, 1999b, *The Venture Capital Cycle* (MIT Press, Cambridge).

[10] Gompers, Paul, and Josh Lerner, 2000, Money chasing deals? The impact of fund inflows on the valuation of private equity investments, *Journal of Financial Economics* 55, 281-325.

[11] Grossman, Sanford, and Oliver Hart, 1982, Corporate financial strucuture and managerial incentives, *The Economics of Information and Uncertainty" (McCall, Ed.)* (Univ. of Chicago Press, Chicago).

[12] Grossman, Sanford, and Oliver Hart, 1983, An analysis of the principal-agent problem, *Econometrica*, 51, 7-45.

[13] Harris, Milton, and Artur Raviv, 1989, The design of securities, *Journal of Financial Economics* 24, 255-287.

[14] Holmström, Bengt, 1979, Moral hazard and observability, *The Bell Journal of Economics* 10, 74-91.

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

[16] Kaplan, Steven N., and Jeremy Stein, 1993, The evolution of buyout pricing and financial structure in the 1980s, *Quarterly Journal of Economics* 108, 313-358.

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

[18] Litvak, Kate, 2009, Venture capital limited partnership agreements: Understanding compensation arrangements, *University of Chicago Law Review* 76, 161-218.

[19] Ljungqvist, Alexander, Matthew Richardson, and Daniel Wolfenzon, 2008, The investment behavior of buyout funds: Theory and evidence, Working paper.

[20] Mirrlees, James A., The optimal structure of incentives and authority within an organization, 1976, *The Bell Journal of Economics*, 7, 105-131.

[21] Nachman, David C., and Thomas H. Noe, 1994, Optimal design of securities under asymmetric information, *Review of Financial Studies* 7, 1-44.

[22] Noe, Thomas H., and Michael J. Rebello, 1992, Adverse selection, contract design, and investment distortion, *Journal of Financial Intermediation* 2, 347-375.

[23] Phalippou, Ludovic, 2009, Beware of venturing into private equity, *Journal of Economic Perspectives* 23, 147-166.

[24] Robinson, David T., and Berk A. Sensoy, 2011, Cyclicality, performance measurement, and cash flow liquidity in private equity, Working paper.

[25] Royden, Halsey L., 1968, *Real Analysis* (MacMillan, New York).

[26] Sahlman, William, 1990, The structure and governnance of venture capital organizations, *Journal of Financial Economics* 27, 473-521.

Essay 2

A Theory of Investment Duration of Venture Capital Funds

2.1 Introduction

By the late 1990s, over 80% of the capital committed to venture capital (VC) funds went to limited partnerships (see Gompers and Lerner 1999). Limited partnerships have a finite lifespan, which is mandated by law in the U.S., and the standard lifespan of a U.S. VC fund is 10-13 years (see Sahlman 1990, Phalippou 2007, and Cumming and Johan 2009). The life of a VC fund can be broadly partitioned into an "investment stage" and a "divestment stage." During the investment stage, a venture capitalist, who serves as the general partner (GP) in the limited partnership, searches for and invests in projects. During the divestment stage, after investments have been brought to fruition, the GP sells the investments of the VC fund in order to realize capital gains.

In the past, the partition of these stages was rather informal: the limited lifespan only imposes an upper bound on the *total* time taken to invest and divest, without further specifying the duration of each stage. However, recently many VC funds have begun to contractually restrict the duration of the investment stage, e.g., the GP cannot make new investments after 5 or 6 years from the start of the partnership (see Kandel, Leschchinskii, and Yuklea 2011). Why do these VC funds impose such a restriction? If such a restriction is beneficial for resolving some problems, why do many other VC funds not implement this restriction?

This paper presents a simple but novel theory that offers an answer to these questions. Basically, we show that when a GP manages a portfolio of projects and has risk-shifting tendencies, the contract that minimizes inefficiency caused by ex post risk-shifting must specify a short investment duration that forces the GP to make his investment decisions on all projects simultaneously. However, short investment duration is not sufficient for minimizing the agency problem; the contract must also satisfy a set of GP's incentive compatibility conditions, which may require the partnership to offer investors rents. When rents offered to investors are too large, the GP, holding all the bargaining power, may prefer a long investment duration financing arrangement, which is rent-saving, at the expense of efficiency.

Below we briefly talk about the model. We model the GP as the agent who has skills in identifying and managing potentially profitable investments but has to rely on external capital provided by investors, who are the limited partners (LPs), to finance these investments. The capital market is supposed to be perfectly competitive. We assume that the GP faces two projects that require financing. Both projects approach the GP simultaneously and both cost the GP only one period from investing to harvesting. After the GP obtains financing, he privately observes the quality of each project, which is unobservable to LPs. Efficiency requires that the GP invests only in positive-NPV (good) projects but not in negative-NPV (bad) ones. We assume that the GP obtains compensation *only if* the VC fund earns a profit.[1] This "limited liability" assumption makes the GP bear little of the downside risk and thus creates an agency conflict.

[1]This assumption is commonly used in the VC literature. See, for example, Axelson, Strömberg, and Weisbach (2009) and Kandel, Leschchinskii, and Yuklea (2011).

We consider three different financing arrangements that the GP could possibly adopt. The first one is *the short (investment) duration capital-unconstrained financing arrangement* (SDCU), under which the GP raises funds for both projects and as the VC fund specifies a short investment duration, the GP has to make investment decisions on both projects simultaneously. The second one we take into account is *the long (investment) duration capital-unconstrained financing arrangement* (LDCU), under which the GP raises funds for both projects and as the VC fund specifies a long investment duration, the GP has the option of investing in one project first and postponing his investment decision on the other project until he observes the outcome of the earlier investment. The third one is *the capital-constrained financing arrangement* (CC), under which the GP raises money sufficient for the investment of only *one* project. Among the three financing arrangements, the GP selects the one that offers him the largest expected payoff.

As LPs are rational, any efficiency loss due to the GP's ex post moral hazard is born by the GP ex ante. Thus, the GP, aiming at maximizing his ex ante expected payoff, has incentives to discipline his ex post investment behavior by ex ante contracting with LPs. Due to limited liability, the GP's risk-shifting tendency can never be perfectly eliminated. Thus, overinvestment can never be completely resolved: it always occurs when both projects are bad, in which case, depending on the contract, the GP either invests in one bad or two bad projects. It is noticeable that by using CC, overinvestment can be kept at its minimum: when there is one good project and one bad project, the GP always invests in the good one; when there are two bad projects, the GP invests in one bad project *only*. However, CC causes underinvestment: when both projects are good, the GP has to pass up one good project because of insufficient funding.

Underinvestment is eliminated when the GP uses a capital-unconstrained financing arrangement. By using such an arrangement, the investment duration specified by the contract becomes important. If the GP uses LDCU, he has an incentive to play *strategic delay* in part of his investments: he has an incentive to invest in the better quality project first and make his investment decision on the poorer quality project contingent on his observation of the outcome of his earlier investment. Note that in our model, the prospects of projects are uncorrelated, so the information on the progress of earlier investments is of no use for the GP to improve the efficiency of later investments. In contrast, such information only helps the GP better exploit the contract to his own benefit at the expense of efficiency. Compared to CC, LDCU causes more overinvestment due to the strategic delay problem: when both projects are bad, the GP not only invests in one bad project but also in the second bad project if the first bad project turns out to be a failure.

SDCU resolves the strategic delay problem, because a short investment duration does not provide the GP with the opportunity to wait and see. We show that, by designing a proper contract under SDCU, the *second best* is attainable. In this second-best world, overinvestment is reduced to the same level as that under CC while underinvestment is completely avoided. Thus, this *second-best arrangement* is more efficient than both CC and LDCU. However, to attain the second best, the contract under SDCU must satisfy a set of GP's incentive compatibility conditions, which may offer LPs positive rents (although the capital market is perfectly competitive). When the rents are large, which happens when the profitability of a successful project is high, the GP may prefer less efficient but rent-saving arrangements, such as CC, when the profitability of a successful project is mildly high, or LDCU, when it is sufficiently high.

Our model suggests a positive correlation between the profitability of a successful project and investment duration of a VC fund. As projects in high-tech industries (biotech, electronics, and Internet) usually own a much higher rate of return if successful compared to projects in traditional industries, we predict that restrictions on investment duration are less likely to take place when the GP specializes in investing in high-tech industries rather than in traditional industries. Moreover, we predict that the recent trend of imposing restrictions on GP's investment duration by partnerships is negatively correlated with project profitability. It would be interesting to test these predictions empirically in the future.

This paper relates to theoretical literature that analyzes the effect of pooling on investment incentives and optimal contracting. Diamond (1984) shows that by changing the cash flow distribution, investment pooling makes it possible to design contracts that incentivize the agent to monitor investments properly. Bolton and Scharfstein (1990) and Laux (2001) show that tying investment decisions together can create "inside wealth" for the agent undertaking investments, which reduces the limited liability constraint and helps design more efficient contracts. Unlike this paper, none of these papers consider the impact of the agent's discretion over the timing of investments on efficiency.

This paper also relates to an emerging literature analyzing private equity fund structures. Kandel, Leschchinskii, and Yuklea (2011) argue that the finite lifespan of a VC partnership can induce a GP to make inefficient investments in risky projects. However, their arguments are based on exogenously specified duration of the VC fund, which has been endogenized in this paper. Inderst, Mueller, and Münnich (2007) argue that pooling private equity investments together in a fund helps a GP commit to efficient liquidation decisions, in a way similar to the winner-picking model of Stein (1997). Their mechanism relies on making the fund capital-constrained, which we show is not always optimal in our model. Axelson, Strömberg, and Weisbach (2009) study the GP's financing of a portfolio of projects. They show that in the pure ex ante financing setting, it is never optimal to make the GP capital-constrained and they find that the optimal financing structure is a combination of ex ante pooled financing and ex post deal-by-deal financing. In contrast, this paper shows that in the pure ex ante financing setting, although pooled financing can increase efficiency, the GP may prefer having the fund capital-constrained.

The remainder of the paper is structured as follows. Section 2.2 outlines the model. Sections 2.3, 2.4, and 2.5 describe the solutions to financing problems under SDCU, LDCU, and CC respectively. Section 2.6 characterizes the equilibrium solution. Section

2.7 concludes. All the proofs are relegated to Appendix.

2.2 The Model

2.2.1 The Basic Framework

There are three types of players in the model: a general partner (GP), investors who are perfectly competitive, and an infinite supply of fly-by-night operators. All players are risk-neutral and have access to a storage technology yielding the risk-free interest rate, which we assume to be zero.

At $t = 1$, two candidate projects arrive. Each requires investment of I to be undertaken. Projects are of two kinds: good (G) and bad (B). The quality of each projects is only observable to the GP at $t = 1$ and is independent of the quality of the other project. Investors only know that with probability λ, a project is good, and with probability $1-\lambda$, a project is bad. A good project has cash flow $R > 0$ with certainty, and a bad project has cash flow 0 with probability $1 - p$ and cash flow R with probability p, where

$$R > I > pR. \tag{2.1}$$

Condition (2.1) implies that good projects are positive NPV investments while bad projects are negative NPV investments. To simplify the discussion, we further assume that

$$R \geq 2I, \tag{2.2}$$

so the costs of investing in both projects can be covered when at least one investment succeeds. Conditions (2.1) and (2.2) imply that

$$p < \frac{1}{2}. \tag{2.3}$$

We assume that a project can be invested either at $t = 1$ or at $t = 2$. However, we assume that for a good project, if it is invested at $t = 2$, its quality turns bad, i.e, its cash flow distribution will be identical to that of a bad project. In contrast to a good project, we assume that a bad project produces the same cash flow distribution no matter whether it is invested at $t = 1$ or at $t = 2$.[2] We assume that a project's cash flow is realized one period after it has been invested, so if the GP invests I in a project at $t = 1$, this project's cash flow is realized at $t = 2$, while if the GP invests I in a project at $t = 2$, this project's cash flow is realized at $t = 3$.

We assume that the GP has no money of his own and finances his investments at $t = 0$, i.e., before the arrival of projects at $t = 1$, by issuing a security $s_I(x)$ backed by the cash flow x from the investments, and keeps the residual security $s_{GP}(x) = x - s_I(x)$.[3] Admissible securities have to satisfy the following limited liability and monotonicity conditions:

Limited Liability (LL): $0 \leq s_I(x) \leq x$,

Monotonicity (M): $s_I(x)$ and $s_{GP}(x)$ are nondecreasing in x.[4]

We assume that proceeds from a project cannot be reinvested in another project. This assumption is a common practice of many VC funds.

Furthermore, we assume that investors cannot distinguish capable GPs from unskilled fly-by-night operators who cannot bring success to any project but can store money at the riskless interest rate. This assumption offers a simple way to rationalize the debt feature

[2]Projects can be considered as business plans in this paper. It is highly plausible that if a good business plan has not been carried out in time, rivals with similar plans will get a head-start in the market, ruining the prospect of an originally good plan. However, for a bad business plan, any similar plans held by rivals are likely to be bad as well, in which case they are less likely to be implemented by rivals. Thus, the assumption that the quality of a bad business plan will not be deteriorated by late implementation is plausible.

[3]Because any project without capital input at $t = 1$ will be a negative-NPV investment at $t = 2$, ex post financing is infeasible.

[4]These assumptions are quite common in the financial literature on security design (see Harris and Raviv 1989; Nachman and Noe 1994). (M) guarantees that the GP cannot increase his own payoff by manipulating the ex post cash flows, either through borrowing money from a third party or burning money.

of GP compensation commonly observed in practice: the GP earns no performance-based compensation before LPs are paid in full.[5] The reason is as follows: if investors are willing to accept a contract that offers the GP a strictly positive payoff by storing money at the riskless interest rate, fly-by-night operators will have incentives to offer investors such a contract. In this case, as the supply of fly-by-night operators is infinitely large, the probability that investors meet a fly-by-night operator will be one. As fly-by-night operators add no value to any project but earn a positive profit, investors must lose money. Thus, any contract that can possibly be accepted by investors must satisfy the following condition.

No Fly-by-night Operators (NF): for invested capital K, $s_{GP}(x) = 0$ whenever $x \leq K$.

Condition (NF) ensures that fly-by-night operators earn no profit from fundraising, so a capable GP is separated from fly-by-night operators by using a contract that satisfies (NF).

2.2.2 Forms of Fundraising

In a first-best world, the GP will invest in all good projects and no bad projects. However, since all securities have to satisfy condition (NF), the GP bears little downside risk. Thus, the GP always has an incentive to undertake some investments rather than undertake no investments. Thus, when both projects are bad, in which case efficiency would require no investments, at least one bad project will be undertaken. Therefore, overinvestment can never be perfectly resolved and the first best is never attainable.

Note that the overinvestment problem is most severe when the two projects are financed separately; in this case, each project's contract has to satisfy (NF), so bad projects will always be undertaken. We will prove later on that it is never optimal to finance the

[5] Admittedly, the assumption that investors cannot distinguish capable GPs from fly-by-night operators may not fit the case of established GPs who have established track records of investments. An alternative way that can overcome this problem is to rationalize the debt component of GP compensation by introducing the effort diversion problem (see Fang 2013). We have not taken this alternative approach as it will add much complexity into the analysis.

two projects separately by looking at three alternative forms of fundraising:

- *The short (investment) duration capital-unconstrained financing arrangement (SDCU).* By using SDCU, the GP raises $2I$ at $t = 0$ to finance both projects jointly and a contract under SDCU specifies a short investment duration so that the GP has to make all investment decisions at $t = 1$; no new investment can be made after $t = 1$.
- *The long (investment) duration capital-unconstrained financing arrangement (LDCU).* By using LDCU, the GP raises $2I$ at $t = 0$ to finance both projects jointly and a contract under LDCU specifies a long investment duration so that investments can either be made at $t = 1$ or $t = 2$.
- *The capital-constrained financing arrangement (CC).* By using CC, the GP raises I at $t = 0$ so that he can only invest in one of the two projects. In this case, the GP will always (weakly) prefer invests I at $t = 1$, so without loss of generality, we assume that a contract under CC specifies a short investment duration.

Below we investigate investment and payoff outcomes under each financing arrangement. Afterwards, we figure out the equilibrium financing arrangement by comparing the GP's expected payoff under each of these three financing arrangements; the one that offers the GP the largest expected payoff will sustain in equilibrium.

2.3 The Short (Investment) Duration Capital Unconstrained Financing Arrangement (SDCU)

By using SDCU, the GP raises $2I$ by issuing $s_I(x)$ to investors at $t = 0$. The final cash flow x can potentially take six different values, i.e., $x \in \{0, I, 2I, R, I + R, 2R\}$. The GP observes the quality of each project at $t = 1$ and decides for each project whether to invest I in the project or store I in a safe asset. By condition (NF), the security design has to satisfy $s_{GP}(2I) = 0$. Thus, when the GP faces two bad projects, he has the incentive to invest in at least one bad project. Although it is impossible to completely contract off overinvestment, the agency conflict can be minimized by using a contract that incentivizes the GP to only invest in good projects when at least one project is good and invest in only one bad project when both projects are bad. We call this financing arrangement *the second-best arrangement*. If the GP uses the second-best arrangement, the GP's maximization problem can be expressed as

$$\max_{s_{GP}(x)} E[s_{GP}(x)] = \lambda^2 s_{GP}(2R) + \left[2\lambda(1-\lambda) + (1-\lambda)^2 p\right] s_{GP}(R+I), \tag{2.4}$$

such that

$$E[x - s_{GP}(x)] \geq 2I \tag{2.5}$$

$$s_{GP}(R+I) \geq p s_{GP}(2R) + (1-p) s_{GP}(R) \tag{2.6}$$

$$p s_{GP}(R+I) \geq p^2 s_{GP}(2R) + 2p(1-p) s_{GP}(R) \tag{2.7}$$

$$s_{GP}(x) = 0 \qquad \forall x \quad s.t.\ x \leq 2I \tag{2.8}$$

$$x - x' \geq s_{GP}(x) - s_{GP}(x') \qquad \forall x, x' \quad s.t.\ x > x'. \tag{2.9}$$

There are two possible payoffs to the GP in the maximand. The first payoff, $s_{GP}(2R)$, occurs only when both projects are good. The second payoff, $s_{GP}(R+I)$, occurs either when only one project is good or when both projects are bad and the invested bad project succeeds.[6]

Condition (2.5) is the investors' break-even condition. The GP has two incentive compatibility conditions, (2.6) and (2.7). The left-hand side of (2.6) is the GP's expected payoff if he only invests in the good project when facing one bad and one good projects. The right-hand side of (2.6) is the GP's expected payoff if he invests in both projects when

[6] When both projects are bad and the invested bad project fails, the final cash flow is I. Due to condition (NF), $s_{GP}(I)$ equals zero and is hence missing in the maximand.

facing one bad and one good projects. When (2.6) holds, the GP only invests in the good project when only one project is good. The left-hand side of (2.7) is the GP's expected payoff if he only invests in one bad project when facing two bad projects. The right-hand side of (2.7) is the GP's expected payoff if he invests in both projects when facing two bad projects. When (2.7) holds, the GP only invests in one bad project when both projects are bad. It is noticeable that when (2.7) is satisfied, (2.6) holds with certainty. Therefore, the GP will follow the prescribed investment behavior if and only if (2.7) is satisfied. Finally, the maximization has to satisfy conditions (NF) and (M). Condition (NF) implies (2.8). Condition (M) can be equivalently expressed into (2.9).

Our first proposition shows that the GP obtains the second-beast surplus by using the second-best arrangement only if the rate of return of a successful project, represented by R/I, is sufficiently low; if otherwise, LPs obtain positive rents, which drives the GP's payoff below the second-best surplus.

PROPOSITION 2.1 *The second-best arrangement is feasible if and only if*

$$\frac{R}{I} \geq \pi_0 \equiv \frac{1+\lambda^2}{2\lambda + (1-\lambda)^2 p}. \tag{2.10}$$

The second-best surplus is

$$[2\lambda + (1-\lambda)^2 p]R - (1+\lambda^2)I, \tag{2.11}$$

which is obtained by subtracting the efficiency loss from undertaking one bad project when both projects are bad from the first-best surplus.

If (2.10) is satisfied and the second-best arrangement is used, the GP obtains the second-best surplus given by (2.11) completely if and only if

$$\frac{R}{I} \leq \hat{\pi}, \tag{2.12}$$

where $\hat{\pi}$ is determined by λ and p. (For the expression for $\hat{\pi}$, see the proof of Proposition 2.1 in Appendix.) If (2.12) is violated, LPs obtain positive rents and the GP's payoff is thus strictly less than the second-best surplus.

To satisfy condition (2.7) so that the GP has an incentive to abandon one bad project when both projects are bad, a contract may give LPs positive rents. This is intuitive. To discourage the GP from gambling on both bad projects, the GP'a payoff when both projects succeed, expressed by $s_{GP}(2R)$, must not be too high compared to his payoff when one bad project succeeds while the other is abandoned, expressed by $s_{GP}(R+I)$. This implies that the GP security cannot pledge too much marginal cash flows in the region $(R+I, 2R]$ to the GP. To satisfy this condition when R/I is high, the contract has to pledge a large fraction of marginal cash flows in the region $(R+I, 2R]$ to LPs, which may offer LPs rents. When such rents are too large, the GP may use other financing arrangements. A question thus occurs: which alternative financing arrangement will the GP use when the second-best arrangement offers LPs too much rents? One candidate could be that the GP still adopts SDCU but uses a contract that violates condition (2.7). In this case, the GP will gamble on two rather than one bad project when both projects are bad. The next proposition is thus straightforward.

PROPOSITION 2.2 *Under SDCU, a contract which violates condition* (2.7) *and satisfies condition* (2.5) *can give the GP an expected payoff weakly less than the following:*

$$2\left[\lambda + (1-\lambda)^2 p\right] R - 2\left[1 - \lambda + \lambda^2\right] I, \tag{2.13}$$

which is obtained by subtracting the efficiency loss from undertaking two bad projects when both projects are bad from the first-best surplus.

Although Proposition 2.2 only gives us an upper bound of the GP's expected payoff by using SDCU with a contract violating (2.7), it will be sufficient for us to show later on that such a financing arrangement will be strictly dominated by LDCU and thus will never be used in equilibrium.

2.4 The Long (Investment) Duration Capital Unconstrained Financing Arrangement (LDCU)

By using LDCU, when both projects are bad, the GP can invest in one bad project at $t = 1$ and make his investment decision on the other bad project at $t = 2$ when the cash flow of the first investment has realized. This strategic delay in part of his investments when both projects are bad offers the GP more information which can be exploited when he makes his second investment decision, so given that the fund offers a long investment duration, this strategy is preferred by the GP. Due to the GP's playing of strategic delay, LDCU results in more overinvestment compared to the second-best arrangement: when both projects are bad, under the second-best arrangement, the GP will invest in only one bad project, while under LDCU, the GP will invest in one bad project at $t = 1$, and if this investment fails, the GP will invest in the second bad project at $t = 2$, since otherwise he obtains nothing.

Although LDCU is less efficient than the second-best arrangement, it is still worth investigating, since the second-best arrangement may offer LPs rents. Note that the maximum efficiency LDCU can achieve occurs when a contract induces the GP to invest only in good projects when at least one project is good and to abandon the second bad project when both projects are bad and the first invested bad project succeeds.

To induce the above prescribed investment behavior, we only need one incentive compatibility condition, which is (2.6). If the GP adopts LDCU, the GP's maximization problem can be expressed as

$$
\begin{aligned}
\max_{s_{GP}(x)} E[s_{GP}(x)] &= \lambda^2 s_{GP}(2R) + \left[2\lambda(1-\lambda) + (1-\lambda)^2 p\right] s_{GP}(R+I) \\
&+ (1-\lambda)^2(1-p)p s_{GP}(R) \qquad (2.14) \\
\text{s.t.} \quad & (2.5),\ (2.6),\ (2.8),\ \text{and } (2.9).
\end{aligned}
$$

There are three possible payoffs to the GP in the maximand. The first payoff, $s_{GP}(2R)$, occurs only when both projects are good. The second payoff, $s_{GP}(R+I)$, occurs either when one project is good or when both projects are bad and the first invested bad project succeeds. The third payoff, $s_{GP}(R)$, occurs when both projects are bad and the first invested bad project fails while the second succeeds.[7]

The following proposition characterizes the optimal contract and the GP's expected payoff under LDCU.

Proposition 2.3 *LDCU is feasible if and only if*

$$
\frac{R}{I} \geq \frac{2 - 2\lambda(1-\lambda) - (1-\lambda)^2 p}{2\lambda + (1-\lambda)^2(2-p)p}. \qquad (2.15)
$$

Under LDCU, an optimal investor security $s_I^L(x)$ (which is not always unique) is given by

$$
s_I^L(x) = \begin{cases} x & \text{if } x \leq 2I \\ 2I & \text{if } 2I < x \leq F_L \\ 2I + x - F_L & \text{if } x > F_L \end{cases}, \qquad (2.16)
$$

where F_L is determined by making investors' break-even condition (2.5) *bind.*

[7]When both projects are bad and both fail, the final cash flow is zero, in which case the GP earns nothing, so this payoff does not enter the maximand.

This security gives LPs no rents. By using LDCU, the GP's expected payoff is

$$\left[2\lambda + (1-\lambda)^2(2-p)p\right] R - \left[2 - 2\lambda(1-\lambda) - (1-\lambda)^2 p\right] I, \tag{2.17}$$

which equals the first-best surplus less the efficiency loss from undertaking one bad project when both projects are bad as well as less the efficiency loss from undertaking the second bad project when both projects are bad and the first invested bad project fails.

Although LDCU is not as efficient as the second-best arrangement, it is obvious that LDCU is more efficient than financing the two projects separately, as LDCU alleviates overinvestment. Moreover, LDCU is more efficient than using a contract violating (2.7) under SDCU. This is because by using a contract that violates (2.7) under SDCU, the GP will invest in two bad projects when both projects are bad, while under LDCU, when both projects are bad, the GP will invest in only one bad project if his first invested bad project succeeds. As by using LDCU, the GP captures all the surplus from investments, LDCU is preferred by the GP to financing the two projects separately and also to using a contract violating (2.7) under SDCU. We enclose this result in the next proposition.

Proposition 2.4 *LDCU is preferred by the GP to financing the two projects separately as well as to using a contract that violates* (2.7) *under SDCU.*

Proposition 2.4 implies that it is never optimal for the GP to finance multiple projects separately. This result is consistent with the practice that the GP finances a portfolio of projects jointly rather than separately. Proposition 2.4 also implies that it is never optimal for the GP to adopt SDCU while at the same time uses a contract that does not sufficiently mitigate overinvestment.

2.5 The Capital-Constrained Financing Arrangement (CC)

Under CC, the GP can only invest in one project between the two. Given the assumptions we made on cash flow distributions of bad and good projects, it is obvious that the GP will always pick up the better quality project to invest in. That is to say, when at least one project is good, the GP will invest in a good project, while when both projects are bad, the GP will invest in a bad project. The next proposition presents the optimal security and the GP's expected payoff following this investment behavior.

Proposition 2.5 *CC is feasible if and only if*

$$\frac{R}{I} \geq \frac{1}{\lambda^2 + 2\lambda(1-\lambda) + (1-\lambda)^2 p}, \tag{2.18}$$

in which case the GP issues debt with face value F_C given by

$$F_C = \frac{I}{\lambda^2 + 2\lambda(1-\lambda) + (1-\lambda)^2 p}. \tag{2.19}$$

This security gives LPs no rents. Under CC, the GP's expected payoff is This payoff can be expressed as follows:

$$\left[\lambda^2 + 2\lambda(1-\lambda) + (1-\lambda)^2 p\right] R - I, \tag{2.20}$$

which equals the first-best surplus less the efficiency loss from undertaking one bad project when both projects are bad as well as less the efficiency loss from missing one good project when both projects are good.

Under CC, overinvestment is minimized; it only occurs when both projects are bad, in which case only one bad project will be invested. Thus, regarding overinvestment mitigation, CC is better than LDCU and is as good as the second-best arrangement. However, the drawback of CC is obvious: when both projects are good, the GP has to miss one good investment opportunity under CC. Thus, CC causes underinvestment, which problem does not occur under LDCU or the second-best arrangement. It is thus clear that CC is less efficient than the second-best arrangement. However, as CC saves rents, CC may be the GP's favorite choice when the second-best arrangement offers LPs too much rents. Compared to LDCU, CC has the advantage of better mitigating overinvestment while the disadvantage of creating underinvestment. It is thus not clear whether the GP will adopt LDCU or CC when the second-best arrangement offers LPs too much rents. In the next section, we derive the GP's optimal financing arrangement by comparing his payoffs from the second-best arrangement, from LDCU, and from CC.

2.6 The Equilibrium Financing Arrangement

Implied by Proposition 2.4, neither financing the two projects separately nor designing a contract that violates condition (2.7) under SDCU will be adopted by the GP in equilibrium. Thus, to find out the equilibrium financing arrangement, we only need to consider the second-best arrangement, LDCU, and CC; the one that gives the GP the largest expected payoff will be the equilibrium financing arrangement. The following theorem gives the result.

THEOREM 2.1 *The equilibrium financing arrangement has the following features:*

(i) As $R/I \to \infty$, the GP either chooses the second-best arrangement or LDCU;

(ii) If the GP chooses the second-best arrangement as $R/I \to \infty$, then he chooses the second-best arrangement for all $R/I \geq \pi_0$, where π_0 is defined by (2.10);

(iii) If the GP chooses LDCU as $R/I \to \infty$, then there must either (a) exist one threshold $\tilde{\pi}$ such that the GP prefers the second-best arrangement when $\pi_0 \leq R/I \leq \tilde{\pi}$ and

prefers LDCU when $R/I > \tilde{\pi}$ or (b) exist two thresholds π_1 and π_2 such that the GP prefers the second-best arrangement when $\pi_0 \leq R/I \leq \pi_1$, prefers CC when $\pi_1 < R/I < \pi_2$, and prefers LDCU when $R/I > \pi_2$.

Remark: Both the second-best arrangement and CC are of short investment duration while LDCU is of long investment duration, so Theorem 2.1 implies that the investment duration of a VC fund is weakly increasing in R/I.

The intuition of Theorem 2.1 is straightforward. The second-best arrangement is more efficient than LDCU and CC, so the second-best arrangement is preferred by the GP when the GP captures all the surplus from investments under this arrangement. As shown by Proposition 2.1, this happens when the rate of return of a successful project, measured by R/I, is low. If R/I is high, it is probable that the second-best arrangement offers LPs too much rents and the GP finds it beneficial to use either LDCU or CC, both of which are rent-saving, at the expense of efficiency. Compared to CC, LDCU has the advantage of eliminating underinvestment while the disadvantage of causing more overinvestment. An increase in R/I increases the efficiency loss from missing a good project (underinvestment) while reduces the efficiency loss from undertaking a bad project (overinvestment). Thus, the GP prefers LDCU to CC when R/I is sufficiently high. This implies that CC could possibly be used only if R/I is neither too high nor too low while LDCU could possibly be used only if R/I is sufficiently high.

By Theorem 2.1, we predict a (weakly) positive correlation between the profitability of a successful project and the investment duration of a VC fund. Note that, whenever a contract specifies a long investment duration, a fund's later investments are on average worse than the fund's earlier investments. We thus predict that, for projects within a VC fund's portfolio, those carried out earlier are more profitable than those invested in later. It would be interesting to test these predictions empirically in the future.

2.7 Conclusion

Investment pooling is a common practice in the VC industry and is considered as a way of mitigating the agency problem. This paper shows that how effective investment pooling alleviates agency conflicts depends on the investment duration specified by a VC fund. If a VC fund specifies a long investment duration, the GP has an incentive to play strategic delay in part of his investments and to make his later investment decisions contingent on the outcomes of his earlier investments. This strategy helps the GP better exploit the contract to his own benefit. To resolve this problem, a VC fund should specify a short investment duration. However, a short investment duration is not sufficient to minimize agency conflicts; the contract must also satisfy a set of the GP's incentive compatibility conditions, which may require the contract to offer LPs rents. When the rents offered to LPs are large, the GP may choose the long investment duration financing arrangement to save rents at the expense of efficiency.

In this paper, we have assumed that the GP holds all the bargaining power, so the financing arrangement is chosen by the GP to maximize his ex ante expected payoff. It would be interesting in the future to study the agency problem analyzed in this paper in the case where LPs hold all the bargaining power. In that case, LPs, just like the GP in this paper, would consider both investment efficiency and rents paid to the other party when maximizing the party's own interests.[8] Would this change in the bargaining power alter the equilibrium financing arrangement and increase investment efficiency? We leave this line of research for the future.

[8]To make the discussion nontrivial, one has to assume that the GP holds a positive reservation value.

2.8 Appendix: Proofs

Outline of proof of Proposition 2.1: Condition (2.7) implies (2.6). Under the second-best arrangement, the GP invests only in good projects when at least one project is good and invests in one bad project when both projects are bad. Such investment behavior leads to the second-best surplus equal to

$$\lambda^2(2R - 2I) + 2\lambda(1-\lambda)(R - I) + (1-\lambda)^2(pR - I). \tag{2.21}$$

Rearranging (2.21) gives (2.11). Since investors must at least break even, (2.21) has to be nonnegative for the GP to raise funds, which requires condition (2.10) to be satisfied.

As the GP security has to satisfy (2.7), it may offer LPs rents and thus the GP may not earn the second-best surplus. Below we solve the program by first solving the program with the investors' break-even condition (2.5) omitted. Lemmas 2.1 and 2.2 offer some characterization of the GP's optimal security that is useful for solving the program. Based on these two lemmas, Lemmas 2.3, 2.4, and 2.5 characterize the GP's maximum expected payoff regardless of the investors' break-even condition. Finally, we take into account the investors' break-even condition, and the GP's actual expected payoff is the minimum between the GP's maximum expected payoff regardless of the investors' break-even condition and the second-best surplus. The expression for $\hat{\pi}$ and the GP's expected payoff when R/I exceeds $\hat{\pi}$ is provided by the proof.

Lemma 2.1 *Holding the expected payment to investors fixed, it is without loss of generality to set $s_{GP}(R+I) - s_{GP}(R)$ as high as possible: either $s_{GP}(R+I) - s_{GP}(R) = I$, $s_{GP}(2R) \geq s_{GP}(R+I)$, and $s_{GP}(R) \geq 0$, or $s_{GP}(R+I) - s_{GP}(R) < I$, $s_{GP}(2R) = s_{GP}(R+I)$, and $s_{GP}(R) = 0$.*

Proof: We first show that given $s_{GP}(R)$, increasing $s_{GP}(R+I)$, which is equivalent to increasing $s_{GP}(R+I) - s_{GP}(R)$, while decreasing $s_{GP}(2R)$ such that the maximand and the expected payment to investors are held constant will relax (2.7). Therefore, as long as conditions (LL), (M), and (NF) are not violated, this will relax the program. Note that when $s_{GP}(R+I) = s_{GP}(R) + I$ (the maximum payment given $s_{GP}(R)$) or when $s_{GP}(2R) = s_{GP}(R+I)$, we cannot do this perturbation without violating (M).

Given $s_{GP}(R)$, increasing $s_{GP}(R+I)$ while decreasing $s_{GP}(2R)$ such that the maximand and the expected payment to investors are held constant requires

$$-ds_{GP}(2R) = \frac{2\lambda(1-\lambda) + (1-\lambda)^2 p}{\lambda^2} ds_{GP}(R+I). \tag{2.22}$$

Obviously, this perturbation relaxes (2.7).

Next it is obvious that decreasing $s_{GP}(R)$ while increasing $s_{GP}(R+I) - s_{GP}(R)$ by the same amount keeps $s_{GP}(R+I)$ and $s_{GP}(2R)$ constant. Hence this perturbation keeps the maximand and the expected payment to investors constant. Note that when $s_{GP}(R) = 0$ or when $s_{GP}(R+I) - s_{GP}(R) = I$, we cannot do this perturbation without violating (M). As this perturbation decreases $s_{GP}(R)$ while keeps $s_{GP}(R+I)$ and $s_{GP}(2R)$ fixed, this perturbation relaxes (2.7). □

LEMMA 2.2 *Given $s_{GP}(R+I) - s_{GP}(R) = I$ and fixing the expected payment to investors, when*

$$p \geq \frac{\lambda}{\sqrt{1 + (1-\lambda)^2} + 1}, \tag{2.23}$$

it is without loss of generality to set $s_{GP}(2R) - s_{GP}(R+I)$ as low as possible: either $s_{GP}(2R) - s_{GP}(R+I) = 0$ and $s_{GP}(R+I) < R - I$, or $s_{GP}(2R) - s_{GP}(R+I) \geq 0$ and $s_{GP}(R+I) = R - I$.

If (2.23) does not hold, it is without loss of generality to set $s_{GP}(2R) - s_{GP}(R+I)$ as high as possible: either $s_{GP}(2R) - s_{GP}(R+I) = R - I$ and $s_{GP}(R+I) \geq I$, or $s_{GP}(2R) - s_{GP}(R+I) < R - I$ and $s_{GP}(R+I) = I$.

Proof: We show that given $s_{GP}(R+I) - s_{GP}(R)$, when (2.23) holds, increasing $s_{GP}(R)$

while decreasing $s_{GP}(2R)$ such that the maximand and the expected payment to investors are held constant will relax (2.7). Therefore, as long as (LL), (M), and (NF) are not violated, this perturbation will relax the program. Note that when $s_{GP}(R) = R - 2I$ or when $s_{GP}(2R) = s_{GP}(R + I)$, we cannot do this perturbation without violating (M).

Given $s_{GP}(R + I) - s_{GP}(R)$, increasing $s_{GP}(R)$ and decreasing $s_{GP}(2R)$ such that the payment to investors and the maximand are held constant will lead to (2.22) and $ds_{GP}(R) = ds_{GP}(R+I)$. Moving all terms in (2.7) to its left-hand side, the change in the left-hand side as we fix $s_{GP}(R + I) - s_{GP}(R)$, increase $s_{GP}(R)$, and decrease $s_{GP}(2R)$ is equal to

$$
\begin{aligned}
& p ds_{GP}(R + I) - p^2 ds_{GP}(2R) - 2p(1 - p)ds_{GP}(R) \\
= \; & p\left[1 + p\frac{2\lambda(1-\lambda) + (1-\lambda)^2 p}{\lambda^2} - 2(1-p)\right] ds_{GP}(R) \\
= \; & \frac{p}{\lambda^2}\left[-\lambda^2 + 2\lambda p + (1-\lambda)^2 p^2\right] ds_{GP}(R). \qquad (2.24)
\end{aligned}
$$

When (2.23) holds, (2.24) is greater than zero if $ds_{GP}(R) > 0$. In other words, when (2.23) holds, the above perturbation relaxes (2.7).

When (2.23) does not hold, (2.24) is greater than zero if $ds_{GP}(R) < 0$. In other words, when (2.23) does not hold, given $s_{GP}(R+I) - s_{GP}(R)$, decreasing $s_{GP}(R)$ while increasing $s_{GP}(2R)$ such that the payment to investors and the maximand are held constant will relax (2.7). Note that when $s_{GP}(R) = 0$ or when $s_{GP}(2R) = s_{GP}(R+I) + R - I$, we cannot do this perturbation without violating (M). $\square$

LEMMA 2.3 *If* (2.23) *holds and* $R/I \leq 3$, *with the investors' break-even condition omitted, the GP security which offers the GP the maximum expected payoff without violating* (2.7) *has* $s_{GP}(R) = R - 2I$, $s_{GP}(R+I) = R - I$, *and* $s_{GP}(2R) = R - I + (3I - R)(1 - p)/p$.

This security offers the GP the expected payoff equal to

$$\left[2\lambda + (1-\lambda)^2 p - \frac{\lambda^2}{p}\right] R - \left[2\lambda + 2\lambda^2 + (1-\lambda)^2 p - \frac{3\lambda^2}{p}\right] I. \tag{2.25}$$

Proof: By Lemmas 2.1 and 2.2, when (2.23) holds, fixing the expected payment to investors, it is without loss of generality to set $s_{GP}(R+I) - s_{GP}(R)$ as high as possible, and given $s_{GP}(R+I) - s_{GP}(R) = I$, to set $s_{GP}(2R) - s_{GP}(R+I)$ as low as possible. In other words, it is without loss of generality to pledge to the GP the cash flows within the region $[R, R+I]$ first, then the cash flows within the region $[2I, R]$, and finally the cash flows within the region $[R+I, 2R]$. When $R/I \leq 3$, it is easy to check that the GP security stated in Lemma 2.3 makes (2.7) bind. Since this security pledges all marginal cash flows within the region $[2I, R+I]$ to the GP, the only way to further increase the GP's maximum expected payoff is by increasing the GP's share of marginal cash flows within the region $[R+I, 2R]$, but this will violate (2.7).

The GP's expected payoff under the second-best arrangement is expressed by (2.4), which equals (2.25) when the GP security takes the form in Lemma 2.3. □

LEMMA 2.4 *If* (2.23) *holds and* $R/I > 3$*, with the investors' break-even condition omitted, the GP security which offers the GP the maximum expected payoff without violating* (2.7) *has* $s_{GP}(R) = I$ *and* $s_{GP}(R+I) = s_{GP}(2R) = 2I$*. This security offers the GP the expected payoff equal to*

$$\left[4\lambda - 2\lambda^2 + 2(1-\lambda)^2 p\right] I. \tag{2.26}$$

Proof: When (2.23) holds, by Lemmas 2.1 and 2.2, it is without loss of generality to pledge to the GP marginal cash flows within the region $[R, R+I]$ first, then marginal cash flows within the region $[2I, R]$, and finally marginal cash flows within the region $[R+I, 2R]$. When $R/I > 3$, it is easy to check that the GP security stated in Lemma 2.4 makes (2.7) bind. Since this security pledges all marginal cash flows within the region

$[R, R+I]$ to the GP and part of marginal cash flows within the region $[2I, R]$ to the GP, if the GP wants to further increase his expected payoff, the GP needs to increase his share of marginal cash flows within the region $[2I, R]$, but this will violate (2.7).

The GP's expected payoff under the second-best arrangement is expressed by (2.4), which equals (2.26) when the GP security takes the form in Lemma 2.4. $\square$

LEMMA 2.5 *If* (2.23) *does not hold and if $R/I > 3 - 2p$, with the investors' break-even condition omitted, the GP security which offers the GP the maximum expected payoff without violating* (2.7) *has $s_{GP}(R) = (I - pR)/(1-p)$, $s_{GP}(R+I) = (I-pR)/(1-p)+I$, and $s_{GP}(2R) = (I - pR)/(1-p) + R$. This security offers the GP the expected payoff equal to*

$$\left[1 + \lambda^2 + (1-\lambda)^2 p - \frac{1}{1-p}\right] R + \left[(1-\lambda)(3\lambda - 1) + (1-\lambda)^2 p + \frac{1}{1-p}\right] I. \tag{2.27}$$

. *If* (2.23) *does not hold and if $R/I \leq 3 - 2p$, the GP security which offers the GP the maximum expected payoff without violating* (2.7) *has $s_{GP}(R) = R - 2I$, $s_{GP}(R+I) = R - I$, and $s_{GP}(2R) = 2R - 2I$. This security offers the GP the expected payoff equal to*

$$\left[2\lambda + (1-\lambda)^2 p\right](R - I). \tag{2.28}$$

Proof: If (2.23) does not hold, by Lemmas 2.1 and 2.2, it is without loss of generality to pledge to the GP marginal cash flows within the region $[R, R+I]$ first, then marginal cash flows within the region $[R+I, 2R]$, and finally marginal cash flows within the region $[2I, R]$. When all marginal cash flows within the region $[R, 2R]$ has been pledged to the GP so that $s_{GP}(R+I) = s_{GP}(R) + I$ and $s_{GP}(2R) = s_{GP}(R) + R$, to have (2.7) satisfied, it is easy to check that $s_{GP}(R)$ must be weakly less than $(I - pR)/(1-p)$. Since $s_{GP}(R)$ cannot exceed $R - 2I$ in order to satisfy (M), when $(I - pR)/(1-p) < R - 2I$, in which case $R/I > 3 - 2p$, the GP's maximum expected payoff without violating (2.7) is achieved when $s_{GP}(R) = (I - pR)/(1 - p)$. In such a case, the GP's expected payoff, expressed by (2.4), equals (2.27). When $(I - pR)/(1 - p) \geq R - 2I$, in which case $R/I \leq 3 - 2p$, the GP's maximum expected payoff without violating (2.7) is achieved when $s_{GP}(R) = R - 2I$. In such a case, the GP's expected payoff, expressed by (2.4), equals (2.28). □

Proof of Proposition 2.1: Lemmas 2.3, 2.4, and 2.5 give the GP's maximum expected payoff without violating (2.7) regardless of the investors' break-even condition. Given the investors' break-even condition satisfied, the GP's expected payoff cannot exceed the second-best surplus given by (2.11). Since the GP maximizes his expected payoff, in each of the situations stated in Lemmas 2.3, 2.4, and 2.5, the GP's expected payoff equals the minimum of the maximum expected payoff stated in the respective lemma and the second-best surplus. The threshold $\hat{\pi}$ can be derived by equalizing the maximum expected payoff and the second-best surplus. In what follows, we display the result without presenting the detailed calculation.

Note that in all the following cases, the GP obtains the second-best surplus expressed by (2.11) if and only if $\pi_0 \leq R/I \leq \hat{\pi}$, where π_0 is defined by (2.10) while the expression for $\hat{\pi}$ varies case by case.

(i) when

$$p \leq \frac{\lambda}{1 + \sqrt{(1-\lambda)^2 + 1}}, \tag{2.29}$$

$\hat{\pi}$ is given by

$$\hat{\pi} = 2 - p + \frac{1 - p}{2\lambda - \lambda^2 + (1-\lambda)^2 p}. \tag{2.30}$$

If R/I exceeds $\hat{\pi}$, the GP's payoff is given by (2.27).

(ii) when

$$p > \max\left\{\frac{\lambda}{1 + \sqrt{(1-\lambda)^2 + 1}}, \frac{1 - 2\lambda - \lambda^2}{(1-\lambda)^2}\right\}, \tag{2.31}$$

$\hat{\pi}$ is given by

$$\hat{\pi} = \frac{3\lambda^2 + (1 - 2\lambda - \lambda^2)p - (1-\lambda)^2 p^2}{\lambda^2}. \tag{2.32}$$

If $\hat{\pi} < R/I \leq 3$, the GP's payoff is given by (2.25), while if $R/I > 3$, the GP's payoff is given by (2.26).

(iii) when

$$\frac{\lambda}{1+\sqrt{(1-\lambda)^2+1}} < p \leq \min\left\{\frac{1}{2}, \frac{1-2\lambda-\lambda^2}{(1-\lambda)^2}\right\}, \tag{2.33}$$

$\hat{\pi}$ is given by

$$\hat{\pi} = 2 + \frac{1-\lambda^2}{2\lambda+(1-\lambda)^2 p}. \tag{2.34}$$

If R/I exceeds $\hat{\pi}$, the GP's payoff is given by (2.26). $\square$

Proof of Proposition 2.2: If a contract violates (2.7), the GP will invest in both projects when both are bad. Given this behavior, efficiency is maximized when the GP invests only in good projects when at least one project is good, in which case the surplus equals

$$\begin{aligned} &\lambda^2(2R-2I) + 2\lambda(1-\lambda)(R-I) \\ &+(1-\lambda)^2\left[p^2(2R-2I) + 2p(1-p)(R-2I) + (1-p)^2(-2I)\right]. \end{aligned} \tag{2.35}$$

Simplifying (2.35) gives (2.13). Since this is the maximum surplus given a contract violates (2.7) and since investors have to at least break even, the GP's expected payoff cannot exceed this surplus. $\square$

Proof of Proposition 2.3: If the GP get financed under LDCU, he will invest only in good projects when at least one project is good, invest only in one bad project when both projects are bad and the first invested bad project succeeds, and invest in both bad projects when both are bad and the first invested bad project fails. Thus, the expected total surplus equals

$$\lambda^2(2R-2I) + 2\lambda(1-\lambda)(R-I) + (1-\lambda)^2(pR-I) + (1-\lambda)^2(1-p)(pR-I), \tag{2.36}$$

which can be simplified into (2.17).

To satisfy the investors' break-even condition (2.5), the surplus expressed by (2.17) must be nonnegative. This requires condition (2.15) to be satisfied.

It is obvious that the investor security s_I^L given by (2.16) satisfies (2.8) and (2.9). We next show that s_I^L always satisfies (2.6). If $F_L \leq R + I$, by (2.16), we know that the GP security has $s_{GP}(R+I) = s_{GP}(2R) \geq s_{GP}(R)$, so (2.6) is satisfied. If $F_L > R+I$, by (2.16), the GP security has $s_{GP}(R) = R-2I$, $s_{GP}(R+I) = R-I$, and $R-I < s_{GP}(2R) < 2R-2I$, so (2.6) still holds.

To maximize his own expected payoff, the GP chooses F_L to make investors just break even. The GP thus captures all the surplus given by (2.17). □

Proof of Proposition 2.4: If the GP finances the two projects separately, by (NF), the GP has an incentive to invest in both projects regardless of their quality. In this case, the surplus from investments will be

$$2\left[\lambda R + (1-\lambda)pR - I\right], \tag{2.37}$$

which is less than the GP's expected payoff under LDCU given by (2.17) under the assumption that $pR < I$. Thus, the GP prefers LDCU to financing both projects separately.

If the GP finances both projects jointly by SDCU with a contract violating (2.7), by Proposition 2.2, his expected payoff cannot exceed that given by (2.13). As (2.13) is less than the GP's expected payoff under LDCU expressed by (2.17) under the assumption that $pR < I$, the GP prefers LDCU to using a contract violating (2.7) under SDCU. □

Proof of Proposition 2.5: The surplus from investing in the better quality project between the two can be expressed as (2.20). To satisfy investors' break-even condition (2.5), (2.20) must be nonnegative, which requires (2.18) to be satisfied.

It can easily be checked that when (2.18) is satisfied, the debt contract with face value given by (2.19) makes investors just break even, in which case the GP captures all the surplus from investments given by (2.20). □

Proof of Theorem 1: To find out the equilibrium financing arrangement, we simply need to compare between the GP's payoff from the second-best arrangement, enclosed in the proof of Proposition 2.1, his payoff from LDCU, presented in Proposition 2.3, and his payoff from CC, presented in Proposition 2.5. The difficulty purely lies in the algebra as the GP's payoff from the second-best arrangement takes different functional forms given different parametric assumptions. In what follows, we present the result without detailed calculation.

- Suppose (2.29) holds.
  - If
    $$\frac{1}{p} \leq 2 - p + \frac{1-p}{2\lambda - \lambda^2 + (1-\lambda)^2 p}, \tag{2.38}$$
    then the second-best arrangement will always be preferred by the GP.
  - If (2.38) does not hold.
    * Suppose
      $$\frac{\lambda^2 + (1-\lambda)^2 - (1-\lambda)^2 p}{\lambda^2 + (1-\lambda)^2 p - (1-\lambda)^2 p^2} \leq \frac{2 + 2\lambda - \lambda^2 - (1 + 2\lambda - \lambda^2)p}{1 - (1-\lambda)^2 (1-p)(1-p+p^2)}, \tag{2.39}$$
      then there exists a threshold $\tilde{\pi}$, expressed by
      $$\tilde{\pi} = \frac{2 + 2\lambda - \lambda^2 - (1 + 2\lambda - \lambda^2)p}{1 - (1-\lambda)^2 (1-p)(1-p+p^2)}, \tag{2.40}$$
      such that the GP prefers the second-best arrangement if $R/I \leq \tilde{\pi}$ while prefers LDCU if otherwise.
    * Suppose (2.39) does not hold, then there exist thresholds π_1 and π_2, expressed by
      $$\pi_1 = \frac{1 + 4\lambda - 3\lambda^2 + (1 - 6\lambda + 4\lambda^2)p - (1-\lambda)^2 p^2}{2\lambda - 2\lambda^2 + (1 - 2\lambda + 2\lambda^2)p} \tag{2.41}$$

$$\pi_2 = \frac{\lambda^2 + (1-\lambda)^2 - (1-\lambda)^2 p}{\lambda^2 + (1-\lambda)^2 p - (1-\lambda)^2 p^2}, \tag{2.42}$$

such that the GP prefers the second-best arrangement if $R/I \leq \pi_1$, prefers CC if $\pi_1 < R/I < \pi_2$, and prefers LDCU if $R/I \geq \pi_2$.

- Suppose (2.31) holds.
  - If
$$\frac{1}{p} \leq \frac{3\lambda^2 + (1 - 2\lambda - \lambda^2)p - (1-\lambda)^2 p^2}{\lambda^2}, \tag{2.43}$$
then the second-best arrangement will always be preferred by the GP.
  - If
$$\frac{3\lambda^2 + (1 - 2\lambda - \lambda^2)p - (1-\lambda)^2 p^2}{\lambda^2} < \frac{1}{p} \leq 3, \tag{2.44}$$
then there exists a threshold $\tilde{\pi}$, expressed by
$$\tilde{\pi} = \frac{2 - 4\lambda - 2(1-\lambda)^2 p + \frac{3\lambda^2}{p}}{(1-\lambda)^2 p + \frac{\lambda^2}{p} - (1-\lambda)^2 p^2}, \tag{2.45}$$
such that the GP prefers the second-best arrangement if $R/I \leq \tilde{\pi}$ while prefers LDCU if otherwise.
  - If $1/p > 3$
    - ∗ Suppose
$$\frac{2 - 4\lambda - 2(1-\lambda)^2 p + \frac{3\lambda^2}{p}}{(1-\lambda)^2 p + \frac{\lambda^2}{p} - (1-\lambda)^2 p^2} \leq 3, \tag{2.46}$$
then there exists a threshold $\tilde{\pi}$, expressed by
$$\tilde{\pi} = \frac{2 - 4\lambda - 2(1-\lambda)^2 p + \frac{3\lambda^2}{p}}{(1-\lambda)^2 p + \frac{\lambda^2}{p} - (1-\lambda)^2 p^2}, \tag{2.47}$$
such that the GP prefers the second-best arrangement if $R/I \leq \tilde{\pi}$ while prefers LDCU if otherwise.

* Suppose (2.46) does not hold, then there exists a threshold $\tilde{\pi}$, expressed by

$$\tilde{\pi} = \frac{2 + 2\lambda + (1-\lambda)^2 p}{2\lambda + 2(1-\lambda)^2 p - (1-\lambda)^2 p^2}, \tag{2.48}$$

such that the GP prefers the second-best arrangement if $R/I \leq \tilde{\pi}$ while prefers LDCU if otherwise.

- Suppose (2.33) holds.
  - If

$$\frac{1}{p} \leq 2 + \frac{1-\lambda^2}{2\lambda + (1-\lambda)^2 p}, \tag{2.49}$$

  then the second-best arrangement will always be preferred by the GP.

  - If (2.49) does not hold,
    * Suppose

$$\frac{2 + 2\lambda + (1-\lambda)^2 p}{2\lambda + 2(1-\lambda)^2 p - (1-\lambda)^2 p^2} \leq \frac{1 + 4\lambda - 2\lambda^2 + 2(1-\lambda)^2 p}{2\lambda - \lambda^2 + (1-\lambda)^2 p}, \tag{2.50}$$

    then there exists a threshold $\tilde{\pi}$, expressed by

$$\tilde{\pi} = \frac{2 + 2\lambda + (1-\lambda)^2 p}{2\lambda + 2(1-\lambda)^2 p - (1-\lambda)^2 p^2}, \tag{2.51}$$

    such that the GP prefers the second-best arrangement if $R/I \leq \tilde{\pi}$ while prefers LDCU if otherwise.

    * Suppose (2.50) does not hold, then there exist thresholds π_1 and π_2, expressed by

$$\pi_1 = \frac{1 + 4\lambda - 2\lambda^2 + 2(1-\lambda)^2 p}{2\lambda - \lambda^2 + (1-\lambda)^2 p} \tag{2.52}$$

$$\pi_2 = \frac{\lambda^2 + (1-\lambda)^2 - (1-\lambda)^2 p}{\lambda^2 + (1-\lambda)^2 p - (1-\lambda)^2 p^2}, \tag{2.53}$$

such that the GP prefers the second-best arrangement if $R/I \leq \pi_1$, prefers CC if $\pi_1 < R/I < \pi_2$, and prefers LDCU if $R/I \geq \pi_2$. $\square$

Bibliography

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

[2] Bolton, Patrick, and David S. Scharfstein, 1990, A theory of predation based on agency problems in financial contracting, *American Economic Review* 80, 93-106.

[3] Cumming, Douglas J., and Sofia A. Johan, 2009, *Venture Capital and Private Equity Contracting: An International Perspective* (Elsevier Science Adademic Press).

[4] Diamond, Douglas, 1984, Financial intermediation and delegated monitoring, *The Review of Economic Studies* 51, 393-414.

[5] Fang, Dawei, 2013, A theory of compensation, performance, and fundraising in the venture capital industry, Working paper.

[6] Gompers, Paul, and Josh Lerner, 1999, An analysis of compensation in the U.S. venture capital partnership, *Journal of Financial Economics* 51, 3-44.

[7] Harris, Milton, and Artur Raviv, 1989, The design of securities, *Journal of Financial Economics* 24, 255-287.

[8] Inderst, Roman, Holger Mueller, and Felix C. Münnich, 2007, Financing a portfolio of projects, *Review of Financial Studies* 20, 1289-1325.

[9] Kandel, Eugene, Dima Leshchinskii, and Harry Yuklea, 2011, VC funds: Aging brings myopia, *Journal of Financial and Quantitative Analysis*, 46, 431-457.

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

[11] Nachman, David C., and Thomas H. Noe, 1994, Optimal design of securities under asymmetric information, *Review of Financial Studies* 7, 1-44.

[12] Phalippou, Ludovic, 2007, Investing in private equity funds: A survey, Working paper.

[13] Sahlman, William, 1990, The structure and governnance of venture capital organizations, *Journal of Financial Economics* 27, 473-521.

[14] Stein, Jeremy, 1997, Internal capital markets and the competition for corporate resources, *Journal of Finance* 52, 111-133.

Essay 3

Skewing the odds: Optimal strategies in capacity-constrained contests (coauthored with Thomas Noe)

3.1 Introduction

3.1.1 Motivation

Contests where parties compete for fixed number of prizes using a fixed bundle of resources are ubiquitous. Consider the problem of a general mapping out strategy for a pivotal battle. The army's capacity, the quantity and quality of the units under the general's command, is fixed. The general's problem is how to deploy this fixed capacity to win a victory. As General Franks said before the first Iraq War, "you have to fight with the army you have." A politician considering how to deploy a fixed campaign war chest over alternative advertising strategies faces a similar problem as does portfolio manager aiming to be ranked as top 10 manager. In an uncertain world, it is natural to represent strategies for such contestants as probability distributions over realized performance and capacities as upper bounds on the expectations of these distributions. Winning a prize depends only on the rank of a contestant's performance relative to the performance of his competitors.

In some cases the capacities of the rival contestants are known. In others, a contestant knows his own capacity but is uncertain about the capacity of the other side. A general might ask, "have the enemy's reinforcements arrived?" The politician, "how much money has my rival raised?" The portfolio manager, "how capable are the other managers?" A natural abstract way to represent this uncertainty is to assume that each contestant has a type, and types are distinguished by their capacities. Each contestant knows his own type but does not know the type of any of the other contestants. The larger the capacity associated with a type, the "stronger" the type. In this paper we consider both the certain and uncertain capacity cases.

The aim of this paper is to investigate these sorts of contests under the assumption of *strategic flexibility:* that is, the only restrictions we impose on the performance distributions contestants may choose are non-negativity and capacity. We ask the following questions: What distributions will the contestants use? How are equilibrium distributions affected by contest parameters? In the case of uncertain capacity, how likely will the "strongest contestants," i.e., the contestants with the greatest capacity, win? How can we improve the power of contests in selecting the strongest contestants?

3.1.2 A Brief Discussion of the Model

Consider two contestants, one weak and the other strong, competing for a single fixed prize. If both contestants chose deterministic performance levels, the strong contestant would always win. However, against the strong contestant choosing a non-stochastic performance level, the weak contestant can win with positive probability by choosing random performance levels. This fact, combined with the insights from Dubins and Savage (1965) on optimal gambling strategies in games with unfair odds, has led many researchers to argue that dispersion is the key parameter in fixed-prize contests and that weak contestants will prefer high dispersion strategies while strong contestants will opt for safety.[1] However, fixed-prize contests are fundamentally different from casino gambles. In contrast to casino gambles, for a contestant in a fixed-prize contest, there is no extra gain from winning big as opposed to merely winning; there is no extra loss from losing big as opposed to merely losing. Given the existence of capacity constraints, this feature of our model makes a win-small/lose-big strategy optimal against any predictable competitor's performance level. In fact, we show that whenever two contestants are equally matched and one of the contestant randomizes performance levels using a strictly unimodal symmetric distribution, the other contestant can always obtain a probability of winning strictly greater than one half by adopting a win-small/lose-big strategy.

Since an optimal best reply for two equally matched contestants will never produce a probability of winning of less than one-half, the win-small/lose-big strategy prevents strictly unimodal distributions from ever being best replies. In the two contestant case, this leads to equilibria in which both contestants submit uniformly distributed performance levels. However, when more than two contestants compete for the prize, uniformly distributed performance levels would generate a probability of winning function that was unimodal which again would be bested by a win-small/lose-big strategy. To prevent this from occurring, the rate of increases individual bidding strategies must decrease, i.e., the bid distribution must become right skewed. This result suggests, and our model verifies, that, typically, equilibrium distributions are highly skewed. At the same time, contestants do not aim to maximize dispersion. Even though there is no exogenous upper bound on the support of the bid distributions, the support is always bounded. Contestants do aim for dispersion locally in that their equilibrium distributions are always absolutely

[1] See, for example, Brown, Harlow, and Starks (1996).

continuous. However global dispersion is limited.

When contestants' capacities are certain and identical, equilibrium distributions are unique. The equilibrium performance level distribution is a complementary beta distribution with one shape parameter being the number of losers and the other being the number of winners. In this case, the results are straightforward. When the contest is selective, i.e., less than one half of the contestants win a prize, equilibrium distributions are right skewed. When the contest is inclusive, i.e., more than one half of the contestants win a prize, equilibrium distributions are left skewed. Equilibrium distributions lack skew only when exactly one half of the contestants win. In this case, equilibrium distributions are symmetric and U-shaped. Equilibrium distributions are never unimodal unless the contest only produces one winner or one loser. When the competition in a contest becomes more intense, either due to an increase in the number of contestants or a decrease in the number of prizes, contestants will increase both the variance and the skewness of their performance distributions, which results in a mean-preserving increase in risk of contestants' performance.

When each contestant is uncertain about the capacities of competitors, the analysis is a bit more involved. However, it is still possible to sharply characterize equilibrium distributions. When strong contestants own sufficiently larger capacity than the weak, the equilibrium is unique. Weak contestants adopt concession strategies that only result in prizes when the realized number of strong contestants is less than the number of prizes. In this case, contestant behavior is not consistent with the risk-taking-and-ruin intuition that agents facing a high probability of loss prefer high variance strategies. In fact, weak contestants' performance distributions may exhibit significantly less dispersion than strong contestants.'[2] When strong contestants are not too strong compared to weak, weak contestants' distributions induce a positive probability of winning even when the

[2]For a discussion of risk taking and the probability of ruin see, for example, Pyle and Turnovsky (1970) and Rose-Ackerman (1991).

realized number of strong contestants exceeds the number of prizes. In this case, there are many equilibria but all equilibria produce the same equilibrium probability of winning conditional on performance. Therefore, even when there are many equilibria, the aggregate outcome is unique and thus it is possible to make sharp predictions about the efficiency of a contest as a selection mechanism.

To this end, we define an efficient contest selection as the one that offers no weak contestant a prize unless all strong contestants receive a prize. It is thus evident, from the above discussion, that contest selection is efficient if strong contestants are sufficiently strong that induces weak contestants to adopt concession strategies in the contest.

As contest selection is not always efficient, whether selection efficiency can be improved by modifying the design of a contest is worth investigating. In this paper, we consider three commonly used modifications: scoring caps, penalty triggers, and localizing contests. The results are clear. Penalizing contestants whose performance fails to reach a threshold improves selection efficiency while dividing a grand contest into smaller local contests harms selection efficiency. Interestingly, capping contestants' performance levels has no impact on selection efficiency (under a weak condition), but it induces contestants to play safer strategies.

3.1.3 Related Literature

This paper contributes to three strands of literature. First, this paper contributes to the growing interest in the impact of tournament incentives on contestants' risk-taking behavior. The existing theoretical literature on this topic often imposes strong restrictions on contestants' risk choices. For instance, some papers assume that a contestant can only choose/mix between a safe and a risky strategy with the performance distribution from the risky strategy exogenously specified (Hvide and Kristiansen 2003, Taylor 2003, and Nieken and Sliwka 2010). Some other papers assume that a contestant's performance distribution is symmetric, such as a normal distribution, and a contestant can only alter the degree of risk-taking by varying the variance of the distribution (Hvide 2002, Kräkel 2008, and Gilpatric 2009). These assumptions have a fundamental impact on contestants' risk-taking behavior: they force the desired tail-risk on one side of the distribution to be accompanied by the undesired tail-risk on the other side and thus force contestants with skewness preference to exhibit variance preference. Indeed, as shown by Gaba, Tsetlin, and Winkler (2004), when contestants can choose any *symmetric* distribution about the same mean, in equilibrium every contestant plays a bang-bang strategy: playing safe when the contest is inclusive while taking extreme risk when the contest is selective. In contrast, we show that neither the safest nor the riskiest strategy would be played in equilibrium if the symmetry assumption is removed and we fully endogenize the equilibrium performance distribution.

Contestants' risk-taking behavior in contests and tournaments has also received interest from empirical studies. One strand of the literature on this topic focuses on mutual-fund managers' risk adjustment according to their relative performance (Brown, Harlow, and Starks 1996, Chevalier and Ellison 1997, and Kempf and Ruenzi 2008). In these studies, risk is typically measured by standard deviations. As shown by this paper, contestants' risk-taking strategies in contests feature skewness, which is supported experimentally by Lin (2011) and Dijk, Holmèn, and Kirchler (2013).[3] This paper thus advocates the incorporation of return skewness into the analysis of mutual-fund tournaments. Under the assumption that the competition between funds is selective, with only a small number of funds identified as "stars" relative to the population of funds, it predicts that the unsystematic returns of delegated fund managers will be right skewed. This prediction is consistent with the empirical evidence Wagner and Winter (2013).

Second, this paper contributes to the literature studying the selection properties of contests.[4] Some existing works take a statistical approach with contestants' performance

[3]Experimental studies on contestants' risk-taking behavior in tournaments only emerge recently. Some other experimental studies include Nieken and Sliwka (2010) and Nieken (2010).

[4]Initiated by the seminal work of Lazear and Rosen (1981), studies of contests as an effort-incentive pro-

distributions exogenously specified (Ryvkin and Ortmann 2008 and Ryvkin 2010) while others take into account contestants' risk-taking (Hvide and Kristiansen 2003) or effort-bidding (Clark and Riis 2001) strategies. In this paper, we study selection efficiency of contests by omitting contestants' effort choices while focusing on their risk-taking behavior, so our approach is closest to Hvide and Kristiansen (2003). Hvide and Kristiansen assume that every contestant can choose between a safe and a risky strategy and they find that an increase in the number of contestants or an increase in the ex ante quality of contestants can sometimes lower the quality of the winner. In contrast, we show that their result no longer holds when contestants are endowed with strategic flexibility in risk-taking.

Finally, this paper contributes to the literature studying strategic resource allocation in competitive environment. Perhaps the most noted problem in this literature is solving the *Colonel Blotto* game: two players simultaneously decide how to assign their "use-it-or-waste-it" resources to different battlefields; the one who assigns more resources to a battlefield wins that battlefield and each player's objective is to win as many battlefields as possible. It is well known that as long as player resources are not very unbalanced, this game has no pure strategy equilibrium.[5] The Colonel Blotto game has an associated General Lotto game where each player, aiming at maximizing the probability of winning, chooses a probability distribution on the nonnegative real line with its mean bounded by the amount of the player's *per battlefield* resources (Hart 2008). General Lotto games can be considered as a relaxed version of Colonel Blotto games because, in General Lotto games, resource constraints only have to be satisfied in expectation, as opposed to realization, as in Colonel Blotto games. As General Lotto games are less technically challenging

vision mechanism are abundant. Among them, a group of papers study the optimal contest/tournament design problem and interestingly, like our paper, some papers find that the grand contest structure is optimal, compared to divisional contest structures (see, for example, Moldovanu and Sela 2006 and Fu and Lu 2009). However, unlike our paper, the optimality of the grand contest structure in these papers is associated with effort-relevant goals, as opposed to our goal of maximizing selection efficiency.

[5]See Roberson (2006) for a characterization of an equilibrium in mixed strategies.

than Colonel Blotto games, General Lotto games are applied more often to economic contexts.[6] This paper, by presenting capacity-constrained contests, extends the General Lotto game from the two-player one-prize case to any n-player m-prize case when players are ex ante symmetric and we also allow each player's resource quantity to be the player's private information, as opposed to public information assumed by the previous models of General Lotto type. As pointed out by Hart (2008), solving the General Lotto game can serve as the first step towards solving the associated Colonel Blotto game.[7] This paper may thus serve as the first step towards solving a more generalized Colonel Blotto game where there are n players rather than two.

3.1.4 Organization of the Paper

The rest of the paper is structured as follows. Section 3.2 analyzes a contestant's optimal distribution choice when his competitor is nonstrategic with a fixed performance distribution. Section 3.3 considers equilibrium solutions when every contestant is strategic with the same fixed capacity. Section 3.4 studies equilibrium solutions when each contestant is uncertain about the capacities of the competitors and presents comparative statics on selection efficiency with respect to contest parameters. Section 3.5 investigates selection efficiency of some modified contest mechanisms. Section 3.6 concludes.

[6] General Lotto games are used in various models of political competition (Myerson 1993, Lizzeri 1999, Lizzeri and Persico 2001, and Sahuguet and Persico 2006).

[7] The second step is to construct a joint resource distribution which satisfies the pointwise resource constraint with each battlefield's marginal resource distribution equal to the distribution derived from the associated General Lotto game (see Hart 2008).

3.2 Besting a Fixed Distribution

3.2.1 Framework

Consider the problem of a contestant picking a distribution function F for a nonnegative random variable, X, so as to maximize his probability of exceeding the realized value of another fixed nonnegative random variable, Y, with distribution P which is statistically independent from X. We call the distribution that the contestant is attempting to surpass the "fixed distribution" and the distribution selected by the contestant the "challenge distribution." To abstract from the problem of ties, in this section, we assume that the fixed distribution is continuous. No continuity restriction is imposed on the challenge variable. Let $F(\cdot)$ and $P(\cdot)$ denote the cumulative distribution functions (CDFs) of the fixed and challenge random variables respectively and let dF and dP represent the measures associated with the random variables. We can express the probability that the challenge variable wins as

$$\mathbb{P}\{X \geq Y\} = \int_0^\infty \mathbb{P}\{Y \leq x\}\, dF(x) = \int_0^\infty P(x)dF(x). \tag{3.1}$$

The contestant's problem is to maximize this probability. It is more convenient to express this problem as the problem of choosing probability measures over the real line rather than random variables on a measure space. Thus, we can formulate the contestant's problem as one of choosing a challenge measure dF to use against the fixed measure dP. The challenge measure has to satisfy two constraints: (a) it has to be a probability measure and (b) its expectation is constrained to be below some value, say $\mu > 0$.[8] We call the latter *the capacity constraint.* We assume that $P(\mu) < 1$. Otherwise the problem is trivial because choosing performance μ will ensure winning with certainty. Since reducing the measure of dF below 1 will never strictly increase the contestant's payoff, the solution to

[8]The mean-constraint assumption is common in the literature studying contestants' risk-taking behavior in contests (e.g., Hvide 2002 and Gaba, Tsetlin, and Winkler 2004.

this problem coincides with the solution to the following relaxed problem:

$$\max_{dF} \int_0^\infty P(x)dF(x) \quad s.t. \int_0^\infty dF(x) \leq 1 \quad \& \int_0^\infty xdF(x) \leq \mu. \tag{3.2}$$

By forming Lagrangian using duality theory we can express the problem as

$$\mathcal{L}(dF, \alpha, \beta) = \int_0^\infty P(x)dF(x) - \alpha \left(\int_0^\infty dF(x) - 1 \right) - \beta \left(\int_0^\infty xdF(x) - \mu \right), \tag{3.3}$$

where α and β are nonnegative dual variables that solve the following dual problem

$$\min_{\alpha,\beta \geq 0} \sup_{dF \in [0,\infty)} \mathcal{L}(dF, \alpha, \beta). \tag{3.4}$$

Equation (3.3) can be rewritten as

$$\mathcal{L}(dF, \alpha, \beta) = \int_0^\infty \left[P(x) - (\alpha + \beta\, x) \right] dF(x) + \alpha + \beta\mu, \tag{3.5}$$

by which it is evident that the optimal dual variables that solve (3.4) must satisfy

$$P(x) - (\alpha + \beta\, x) \leq 0, \quad \forall\, x \geq 0, \tag{3.6}$$

since otherwise $\sup_{dF \in [0,\infty)} \mathcal{L}(dF, \alpha, \beta)$ goes to positive infinity. Thus, $\alpha + \beta\, x$ is an upper bound for $P(x)$. When condition (3.6) is satisfied, the optimal value of the dual problem equals $\alpha + \beta\mu$, which is strictly increasing in both α and β. Thus, given condition (3.6) satisfied and $\alpha, \beta \geq 0$, to solve the dual problem, α and β must both be set as low as possible. Therefore, condition (3.6) must be binding for some $x \geq 0$, i.e, there exists some point(s) $x' \geq 0$ such that $P(x') - (\alpha + \beta\, x') = 0$. Thus, $\alpha + \beta\, x$ is an upper support line for P. Placing any mass on points at which $P(x) - (\alpha + \beta\, x) < 0$ lowers the Lagrangian. Thus the optimal challenge measure places no weight on such points and is always concentrated on an upper support line, $\alpha + \beta\, x$, for $P(\cdot)$. Therefore, the optimal challenge measure,

dF, together with the optimal dual variables, α and β, must satisfy the following:

$$\begin{aligned} P(x) &\leq \alpha + \beta\, x \quad \forall\, x \geq 0 \\ dF\{x \geq 0 : P(x) &< \alpha + \beta\, x\} = 0. \end{aligned} \tag{3.7}$$

3.2.2 Besting Specific Distributions

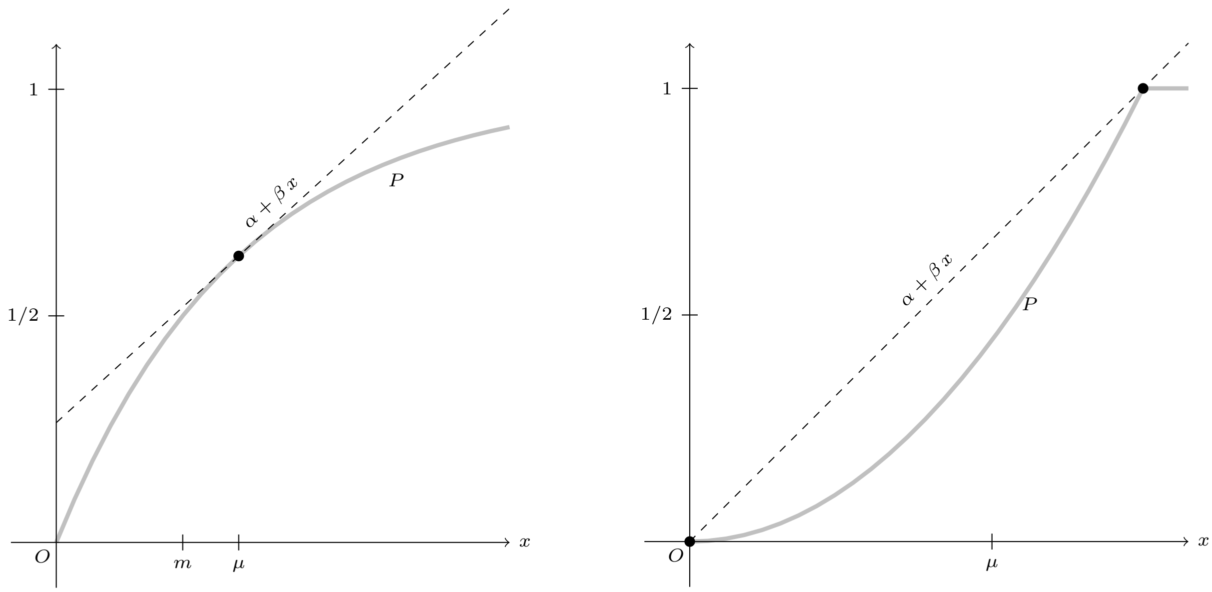
The optimal challenge measure depends on the shape of $P(\cdot)$. If $P(\cdot)$ is strictly concave on its support, then at each point on the graph of $P(\cdot)$, there is an upper support line that is strictly above $P(\cdot)$ at all other points. This implies that the optimal challenge distribution places all weight on a single point. If $P(\cdot)$ is strictly convex on its support, its support must be bounded in order for P to be a CDF. Assume that the support is $[0, \zeta]$. Recall that we have assumed $P(\mu) < 1$ in order to make the discussion nontrivial. This assumption is equivalent to assuming $\zeta > \mu$ here. Note that when $P(\cdot)$ is strictly convex on its support $[0, \zeta]$, the upper support line that intersects the graph of $P(\cdot)$ must intersects P either (i) *only* at zero, or (ii) *only* at ζ, or (iii) at both zero and ζ. As placing all weight at zero is obviously not optimal while placing all weight on ζ violates the capacity constraint, the upper support lines that satisfy (i) or (ii) cannot be the support lines associated with the optimal challenge measure. Therefore, the only upper support line associated with the optimal challenge measure must be the one that satisfies (iii); it must connect the value of $P(\cdot)$ at 0 to the value of $P(\cdot)$ at its upper endpoint, ζ, so the optimal measure only places weight on 0 and ζ, the lower and upper endpoints of the support.

Suppose that the capacity constraint, μ, equals the mean of P. In this case, we can interpret our problem as one picking a distribution, F, with the best chance of besting a fixed distribution P with the same mean. One possible solution is to set $F = P$ which would yield a probability of winning equal to 1/2. Is it possible to do better, i.e., can an contestant "best" P by garnering a probability of winning exceeding 1/2?

First, consider the case where P is strictly concave. Consider the optimal distribution against a given concave fixed distribution detailed above. This distribution places all probability weight on a single point. Because the expectation of P equals μ, the play-safe distribution places all probability weight on the point μ. This distribution yields a probability of winning equal to $P(\mu)$. By Jensen's inequality, since μ is the mean of the random variable, Y, and P is concave, $P(\mu) > \mathbb{E}[P(Y)]$. Because P is the distribution of Y, $\mathbb{E}[P(Y)] = 1/2$ and thus, the payoff from the play-safe distribution $P(\mu)$ exceeds $1/2$. Thus, "playing safe" always bests a strictly concave CDF with the same mean. This result is illustrated in Panel A of Figure 3.1.

Now consider the case where P is convex with support $[0, \zeta]$. As shown earlier, optimal challenge distribution to play against P involves placing all probability mass on the points 0 and ζ. In order for the challenge distribution to have expectation μ, the weight on ζ, which also equals the probability of the challenge distribution winning, must equal μ/ζ. Because P is strictly convex and supported by $[0, \zeta]$, P stochastically dominates the uniform distribution over $[0, \zeta]$. Hence, the mean of P exceeds the mean of the uniform distribution, i.e., $\mu > \zeta/2$. Thus, the probability of winning using the challenge distribution, μ/ζ, exceeds $1/2$. Therefore, the "playing risky" distribution which places all weight on the extreme points always bests a strictly convex CDF with the same mean. This result is illustrated in Panel B of Figure 3.1.

However, a strictly convex CDF implies an increasing probability density function (PDF). This is not a common property for "textbook" distributions. Unimodal symmetric distribution functions are far more commonly encountered in the economics and statistics literature. Is it possible to best these distributions? In fact, the optimal challenge distribution against such distributions is a "win-small/lose-big" distribution. Suppose that P is a strictly unimodal symmetric distribution. Let x^* be defined as the maximizer of $P(x)/x$ over the support of P. x^* exists and exceeds the mode of the distribution. By symmetry, this implies that x^* exceeds the mean of the distribution as well. Consider



Panel A. Concave case

Panel B. Convex case

Figure 3.1: Besting convex and concave CDFs. The graphs illustrate optimal challenge distributions played against fixed distributions whose mean equals the mean of the challenge distribution. The upper support lines are represented by dashed lines. The fixed distributions are represented by thick grey lines. The expectations and medians of the fixed distributions are represented by μ and m respectively. Panel A represents the case where the fixed distribution is concave and Panel B represents the case where the fixed distribution is convex.

the line connecting the origin with $(x^*, P(x^*))$. This line is an upper support line for P. Consider a distribution that places weight of μ/x^* on x^* and weight of $(1 - \mu/x^*)$ on 0. Then the expectation of this distribution is μ and the probability of winning is

$$(\mu/x^*)\, P(x^*). \tag{3.8}$$

Note that by the definition of x^*,

$$\frac{P(x^*)}{x^*} > \frac{P(x)}{x} \quad \forall\, x \neq x^*. \tag{3.9}$$

Thus,

$$P(x^*)\, x > x^*\, P(x). \tag{3.10}$$

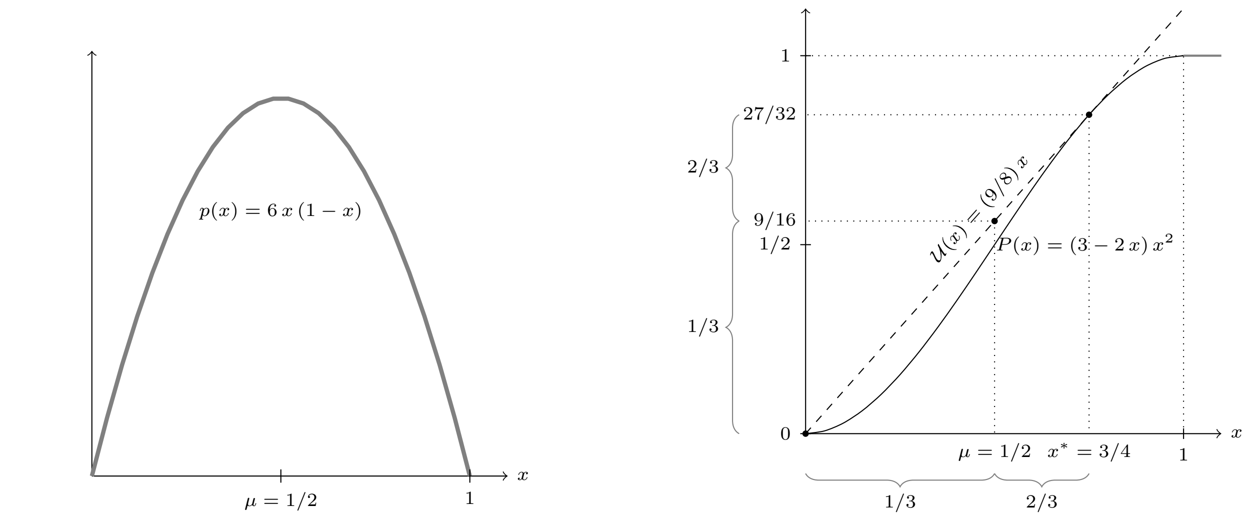
Integrating both sides over dP yields

$$P(x^*)\,\mu = P(x^*) \int x\,dP(x) > x^* \int P(x)dP(x) = x^*\,\frac{1}{2}. \tag{3.11}$$

Combining (3.8) and (3.11) shows that a symmetric strictly unimodal fixed distribution can always be bested by the win-small/lose-big challenge distribution. The construction of the optimal challenge distribution is illustrated by Figure 3.2. In fact, the win-small/lose-big distribution bests asymmetric unimodal distributions whenever the fixed distribution's mean, μ, is less than x^*. A sufficient condition for $\mu < x^*$ is for the mean-median-mode inequality to hold. If the opposite inequality holds, the mode-median-mean inequality, then $1/2 = P(\text{median}) < P(\mu)$ and thus a pay-safe distribution bests the unimodal distribution. The "typical" case for unimodal distributions is for either the mean-median-mode or the mode-median-mean inequality to hold, and there are a number of results in the statistical literature identifying sufficient conditions for one of these inequalities to hold.[9] Thus, typically a unimodal distribution can be bested by a simple distribution, either a win-small/lose-big or a play-safe distribution.

These specific results can be extended to find challenge distributions that best many other classes of fixed distributions; however, it is impossible to enumerate all of them. Nevertheless, we can answer one more general question. Is there a distribution that cannot be bested? The best reply to such a distribution would have to yield a probability of winning equal to 1/2. This is the probability of winning obtained by simply using challenge distribution P itself against the fixed distribution P. Thus, a distribution that cannot be bested must be the one which is a best reply to itself. One such distribution is the uniform distribution over $[0, 2\,\mu]$. To see this, note that the probability of winning against a uniform distribution with mean μ is $P(x) = x/(2\mu)$. Thus for any distribution,

[9] See Sato (1997) for further discussion of mode-median-mean inequality.



Panel A. PDF Panel B. CDF

Figure 3.2: Panel A illustrates the probability density function, p, for a symmetric unimodal distribution — the beta distribution with shape parameters $(2, 2)$. Panel B illustrates a distribution which bests the beta distribution: The CDF, P, of the beta distribution is represented by the grey line; the upper support line, $\mathcal{U}$, for the CDF by the dashed line. The distribution which bests this beta distribution places all its weight on two points, $x = 0$ and $x^* = 3/4$, where the support line meets the distribution function. The probabilities of these points are set so that the expected value of the challenge distribution equals the expected value of the beta distribution, Thus, the probability of $x = 0$ equals 1 and the probability of $x = x^* = 3/4$ equals $2/3$. The probability that the challenge distribution will win is given by the probability that $x = x^* = 3/4$ times the probability of winning when $x = x^*$, $P(x^*) = 27/32$. Thus, the probability that the challenge distribution wins is $(2/3)(27/32) = 9/16 > 1/2$.

F over the support of P with mean μ

$$\int P(x)dF(x) = \int \frac{x}{2\mu} dF(x) = \frac{1}{2\mu} \int x\, dF(x) = \frac{1}{2}. \tag{3.12}$$

Is this the only distribution that is a best reply to itself? To answer this question we require the following technical lemma.

Lemma 3.1 *Suppose dP is a finite measure over $[0, c)$. Let P be the distribution function associated with dP. Let $\bar{P}$ be an non-decreasing absolutely continuous function defined on $[0, c)$ and let $d\bar{P}$ be the associated measure. Suppose that* *(i) For all* $x \in [0, c)$, $P(x) \leq \bar{P}(x)$

(ii) For some $x' \in [0, c)$, $P(x') < \bar{P}(x')$

(iii) And for every measurable set, $\mathcal{E}$,

$$dP\{x \in [0, c) : P(x) < \bar{P}(x)\} = 0. \tag{3.13}$$

Then P *is absolutely continuous and there exist* $\hat{x} < c$ *and* $p^* < \infty$*, such that*

$$P(x) = \begin{cases} \bar{P}(x) & x \leq \hat{x} \\ p^* & x \in [\hat{x}, c) \end{cases} \tag{3.14}$$

Proof: See Appendix. $\square$

The following basic result follows directly from Lemma 3.1.

Proposition 3.1 *If a distribution* P *has the property that the maximum probability of winning against* P *for all challenge distributions with the same mean equals 1/2, then* P *is a uniform distribution.*

Proof: If the maximum for problem 3.5 equals 1/2 then P itself must be a maximizer for problem 3.5 as playing P against P always yields 1/2. This implies that the necessary conditions (3.7) are satisfied with $F = P$, i.e.

$$P(x) \leq \alpha + \beta\, x \quad \forall\, x \geq 0 \tag{3.15}$$

$$dP\{x \geq 0 : P(x) < \alpha + \beta\, x\} = 0. \tag{3.16}$$

By Lemma 3.1, this implies that

$$P(x) = \begin{cases} \alpha + \beta\, x & x \leq \hat{x} \\ 1 & x > \hat{x} \end{cases} \tag{3.17}$$

The continuity of P at 0 implies that $\alpha = 0$. Thus, P is a uniform distribution. $\square$

3.3 Contests without Capacity Uncertainty

3.3.1 Probability of Winning

In the previous section, we studied a contestant's optimal distribution choice under a decision-theoretic framework in which the competitor's distribution choice is fixed. In this section, we introduce the contest game in which every contestant acts strategically in order to win the contest.

The contest involves a total number of n contestants competing for m prizes. A contestant can win at most one prize. For analytical convenience, we assume that each prize has the same positive value. Each contestant simultaneously and effortlessly chooses a probability distribution from which his realized performance is drawn. Distributions chosen by different contestants are assumed to be independent. Afterward, each contestant's performance level is realized and ranked. Prizes are then distributed to the contestants according to their rankings from top to bottom until all the m prizes are distributed. Tied contestants have the same chance of winning a prize. All the losers receive a payoff equal to 0.

We impose two conditions on admissible distribution choices of a contestant. First, the support of the distribution must be on the nonnegative real line. Second, the expected performance, i.e., the mean of the performance distribution, cannot exceed the contestant's capacity. In this section, we assume that all contestants have the same capacity equal to μ. This assumption will be relaxed in the next section. Given these two conditions, we aim to figure out the equilibrium distribution chosen by each contestant. The case where $m = 0$ and the case where $m = n$ are trivial as in both cases a contestant's performance has no impact on his payoff and thus any distribution is optimal, so in the rest of the paper, unless we explicitly present the constraints on m and n, we take for granted that $1 \leq m \leq n - 1$.

We focus on symmetric Nash equilibria throughout the paper, in which contestants with the same capacity play the same distribution. In this section, this means that all contestants play the same distribution as they have the same capacity. Because the contestants' choice sets are convex, we do not need to consider mixed strategies. By symmetry, the probability of winning function, P, is the same for all contestants. Contestant i's performance is a random variable, denoted by X_i, and his realized performance is x_i, which is drawn from the distribution he chooses, denoted by $F_i(\cdot)$. Because we concentrate on symmetric equilibria and all contestants are homogeneous, we suppress the index of a contestant's identity. In order to win a prize, a contestant's realized performance, x, has to be at least the mth-largest among all the contestants. Thus a contestant's probability of winning, denoted by $P(x)$, is the probability that the contestant's performance tops the mth best performance of the remaining $n - 1$ contestants. Given symmetry, this probability of winning, P, is given by the cumulative distribution of an order statistics produced by the equilibrium distribution, F. We will talk about this in detail later. However, since a contestant is only affected by other contestants' performance distributions through the probability of winning function, P, we can partially characterize equilibrium behavior without making explicit reference to the relation between P and F. Our first such characterization shows that in any symmetric Nash equilibrium, P is continuous and intersects the origin.

LEMMA 3.2 *If P is the probability of winning function produced by a symmetric Nash* *equilibrium, then P is continuous and $P(0) = 0$.*

Proof: If P is not continuous, then it must have a point of discontinuity. Call this point $x^o \geq 0$. The discontinuity at x^o implies positive mass at x^o. Thus, x^o must be a best reply for every contestant. However, this is not possible because a contestant is always better off transferring mass from x^o to $x^o + \epsilon$, for ϵ sufficiently small. Such a transfer's effect on the capacity constraint can be made arbitrarily small by shrinking ϵ to zero while for all positive ϵ, no matter how small, the transfer generates a gain that is bounded by a strictly positive number from below. As no contestant places any point mass in equilibrium, $F(0) = 0$. As P is the CDF of order statistics of F, $F(0) = 0$ implies $P(0) = 0$. $\square$

Thus, P satisfies the conditions for the maximization problem in Section 3.2, so the distribution F played by each contestant against P together with the two nonnegative Lagrange multipliers α and β must satisfy (3.7). Hence P satisfies the conditions of Lemma 3.1 with $\bar{P}(x)$ replaced by $\alpha + \beta\, x$. Applying Lemma 3.1 gives the following result.

PROPOSITION 3.2 *If every contestant has the same fixed capacity, there exist nonnegative constants, α, β, and $\hat{x}$, such that if P represents the probability of winning for a contestant in a symmetric Nash equilibrium,*

$$P(x) = \begin{cases} \alpha + \beta\, x & \text{if } x \leq \hat{x} \\ 1 & \text{if } x > \hat{x} \end{cases} \tag{3.18}$$

where $\hat{x}$ is defined by

$$\alpha + \beta\, \hat{x} = 1. \tag{3.19}$$

Proposition 3.2 implies that the support of F is an interval $[0, \hat{x}]$. In this interval, P is linear and equals the (identical) support lines of each contestant. Next, we need to characterize the interval of the common support of P and F and the multipliers for the contestant's optimization problem. The next lemma provides this characterization.

LEMMA 3.3 *The Lagrange multipliers from the dual problem are:* $\alpha = 0$ *and* $\beta = \frac{m}{n\mu}$*. The upper bound of the support of* F *is* $\hat{x} = \frac{n\mu}{m}$*.*

Proof: By Lemma 3.2, $P(0) = 0$. Thus, by (3.18), $\alpha = 0$ and $P(x) = \beta x$ on its support. As P is linear on its support, it is weakly optimal for a contestant to play safe. Thus a contestant's probability of winning equals $P(\mu) = \beta\mu$. As there are m prizes for n contestants, by symmetry, each contestant has probability m/n to win. Thus, $\beta\mu = \frac{m}{n}$, so $\beta = \frac{m}{n\mu}$. By (3.19), $\hat{x} = \frac{n\mu}{m}$. $\square$

3.3.2 Individual Contestant Strategies

By Proposition 3.2 and Lemma 3.3, the probability of winning function, P, is fully characterized. To solve for the equilibrium strategy, F, we only need to work out the relation between P and F.

Consider a contestant with a realized performance level equal to x. He has $(n-1)$ opponents, each choosing an identical and independent distribution F. As no one places point mass, F is continuous, so there is no chance of a tie. His probability of winning a prize equals the probability that his realized performance exceeds mth highest performance of remaining $n-1$ contestants. Note that mth highest performance out of $n-1$ is the $(n-m)$th lowest performance. Thus the contestant's probability of winning given performance x, represented by $P(x)$, equals

$$\mathbb{P}[X_{n-m:n-1} \leq x], \tag{3.20}$$

where $X_{n-m:n-1}$ represents the $(n-m)$th order statistic for the random variable X with distribution F. Thus, the probability of winning function, P, equals the distribution of this order statistic, which is given by (see Lemma 1.3.1 in Reiss (1980))

$$P(x) = F_{n-m:n-1}(x) = \sum_{i=n-m}^{n-1} \binom{n-1}{i} F(x)^i \, (1 - F(x))^{(n-1)-i}. \tag{3.21}$$

As P is fully characterized by Proposition 3.2 and Lemma 3.3, the following proposition which characterizes the equilibrium distribution is straightforward by equation (3.21).

Proposition 3.3 *In a contest where there are n contestants with the same fixed capacity, μ, competing for m prizes, there exists a unique symmetric Nash equilibrium in which each contestant picks a distribution F that satisfies the following equation on its support $[0, \frac{n\mu}{m}]$:*

$$\sum_{i=n-m}^{n-1} \binom{n-1}{i} F(x)^i \, (1 - F(x))^{(n-1)-i} = \frac{m}{n\mu} x. \tag{3.22}$$

By Proposition 3.3, when $n = 2$ and $m = 1$, $F(x) = \frac{x}{2\mu}$ on the support $[0, 2\mu]$, so F is a uniform distribution in this case, confirming our finding in Proposition 3.1. When $m = 1$ and $n > 2$, F is a power-function distribution with CDF given by $F(x) = \left(\frac{x}{n\mu}\right)^{\frac{1}{n-1}}$ on the support $[0, n\mu]$. It seems that F belongs to different types of distributions when n or m changes. Is there a single category of distribution which F belongs to in general when F satisfies equation (3.22)? The answer is yes! In fact F is a stretched *complementary beta distribution*, which coincides with some other types of distributions in some special cases. The complementary beta distribution was first named and studied by Jones (2002). It is obtained by swapping the roles of the CDF and the quantile function for the standard beta distribution. Jones defines the complementary beta distribution in a general way where the two shape parameters characterizing the beta distribution can be any positive numbers. As it turns out that we only need to look at the cases where the two shape parameters are positive integers, we use a confined definition.

It is well known that if a random variable Y has a standard beta distribution on the support $[0, 1]$ with two shape parameters being two positive integers, a and b, then its

CDF, F_Y can be expressed as

$$F_Y(y) = \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} y^j (1-y)^{a+b-1-j}. \tag{3.23}$$

Let $u = F_Y(y)$, so $F_Y^{-1}(u)$ is the quantile function of the standard beta distribution. If the CDF of a random variable U on the support $[0,1]$ satisfies $F_U(u) = F_Y^{-1}(u)$, then U has a complementary beta distribution with shape parameters being a and b. We present the concise definition below.

Definition 3.1 *If U is a random variable on the support $[0,1]$ whose CDF, F_U, satisfies*

$$\sum_{j=a}^{a+b-1} \binom{a+b-1}{j} F_U(u)^j (1-F_U(u))^{a+b-1-j} = u, \tag{3.24}$$

then U has a complementary beta distribution with shape parameters being a and b, denoted by $U \sim CB(a,b)$.

If we scale the random variable X whose CDF satisfies (3.22) by $\frac{n\mu}{m}$ to create another random variable Z such that $Z \equiv \frac{n\mu}{m}X$, then the CDF of Z satisfies

$$\sum_{j=n-m}^{n-1} \binom{n-1}{j} F_Z(z)^j (1-F_Z(z))^{n-1-j} = z. \tag{3.25}$$

Note that if $a = n-m$ and $b = m$, equation (3.24) coincides with equation (3.25), so by definition, $Z \sim CB(n-m,m)$. Thus $\frac{m}{n\mu}X \sim CB(n-m,m)$, implying that a contestant's equilibrium performance is complementary-beta distributed. The next proposition summarizes our observation.

Proposition 3.4 *If X represents a contestant's random performance, then in a symmetric Nash equilibrium, $\frac{m}{n\mu}X \sim CB(n-m,m)$.*

Note that the two shape parameters of the equilibrium distribution, $(n-m)$ and m, coincide with the number of losers and the number of winners respectively in our model. Jones (2002) shows that the PDF of the complementary beta distribution is convex and U-shaped with a unique antimode when the two shape parameters are strictly greater than 1, i.e., when $1 < m < n-1$ in our case. The PDF is unimodal only when the contest is either a one-winner contest ($m = 1$) or a one-loser contest ($m = n-1$).[10] The existing literature studying contestants' risk-taking behavior in contests usually assumes unimodality of admissible distributions by assuming, for example, normal distributions (e.g., Taylor 2003 and Nieken and Sliwka 2010). Our results suggest that the exogenously specified symmetric unimodal distributions are the antithesis of the distributions that emerge endogenously in equilibrium.

Gaba, Tsetlin, and Winkler (2004) demonstrate that, when contestants are only allowed to choose among distributions whose PDFs are *symmetric* about the same value, the weakly dominant strategy for each contestant is a bang-bang strategy: choose the most risky distribution if the contest is "selective", i.e., if $m/n < 1/2$, and the least risky one if the contest is "inclusive", i.e., if $m/n > 1/2$. In contrast, we find that, if the symmetry assumption is relaxed, the bang-bang strategy is never the equilibrium strategy; instead, in equilibrium, contestants disperse probability measures over a continuous and finite interval that creates uncertainty in their performance levels but performance dispersion is limited. This result may shed some light on the findings from Falkenstein (1996) who, through his empirical study of mutual fund portfolio holdings, reveals tournament incentives owned by mutual funds and suggests that mutual funds do not seem to prefer the most highly volatile stocks but they seem to avoid the lowest volatility stocks.

In the following subsections, we apply the properties of the complementary beta distribution studied by Jones (2002) to examine the impact of a change in n, m, or μ on the shape of the equilibrium distribution, F, focusing on the changes in its L-scale and

[10]When $m = 1$ and $n = 2$, the distribution is uniform; in all the other one-winner or one-loser contests, the PDF is convex and is both unimodal and uniantimodal.

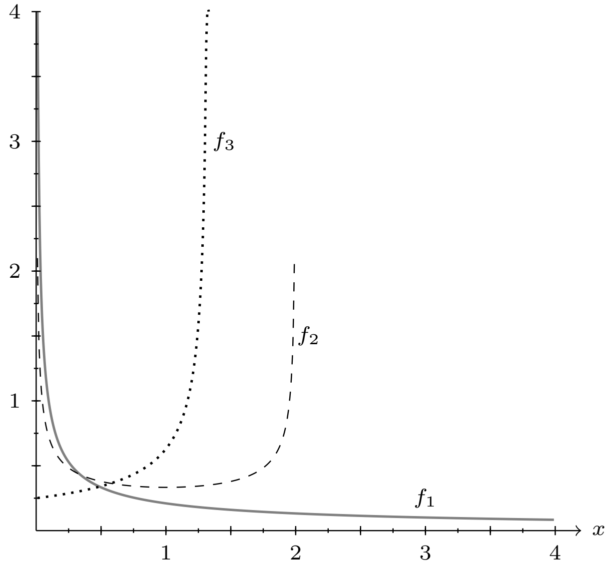
L-skewness. The L-scale is the second L-moment of a distribution and equals half of Gini's mean difference. It is a measure of the scale or dispersion of the distribution analogous to standard deviation. Both the L-scale and the standard deviation satisfy the conditions specified in Oja (1981) for dispersion measures. The L-skewness is the third L-moment ratio, calculated as dividing the third L-moment by the L-scale. It is location and scale invariant and is a measure of the asymmetry of a distribution analogous to the conventional skewness (see Hosking 1990). Both the L-skewness and the conventional skewness measure satisfy the conditions specified in Oja (1981) for skewness measures. We examine the L-scale and the L-skewness instead of the conventional standard deviation and skewness because the complementary beta distribution is not well suited to the explicit computation of moments but well suited to the explicit computation of the L-moments, which are expectations of certain linear combinations of order statistics (see Jones 2002).

3.3.3 Selectivity

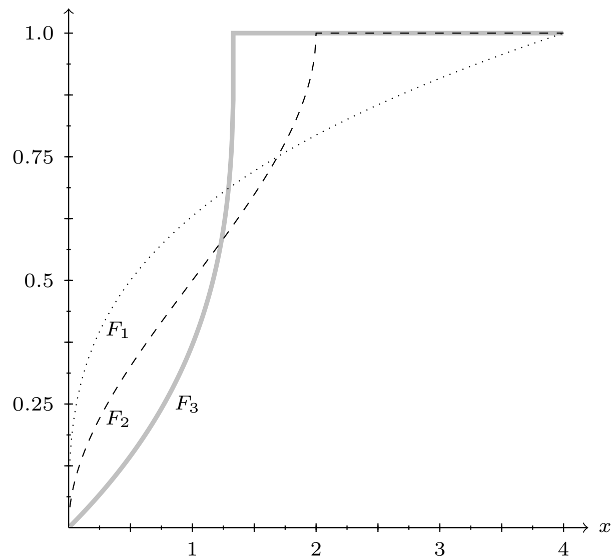
By Proposition 3.4, each contestant chooses a stretched $CB(n-m, m)$. Making the contest more selective increases the right skewness of the equilibrium distribution. The intuition for this result is easiest to understand if we restrict attention to the contest with fixed number of contestants and we increase selectivity by decreasing the number of prizes. As the number of prizes falls, for any fixed distribution selected by the contestants, the probability of a given contestant winning over the high performance level range relative to the low performance level range increases.[11] As shown in Proposition 3.2, in equilibrium, the marginal incentives must be the same at all performance levels in the support of the equilibrium distribution. Thus, the equilibrium distribution function's slope at the high end must decrease relative to the low end to compensate, i.e., skew must increase.

This result is illustrated in Figure 3.3 and demonstrated in Proposition 3.5.

[11] This effect of selectivity on the relative likelihood of the high versus low range follows because the distribution of marginal winning bid in the more selective contest dominates the distribution of marginal winning bid in the less selective contest by first-order stochastic dominance (Nanda and Shaked 2001).



Panel A. PDF



Panel B. CDF

Figure 3.3: *Effect of selectivity on the equilibrium distributions.* The figure plots equilibrium distributions for contests with four contestants. f_i (F_i) represents the equilibrium PDF (CDF) when i winners are selected. Increasing selectivity increases the right skewness and dispersion of the equilibrium distribution.

Using the properties of the complementary beta distribution, we characterize the effects of selectivity on dispersion and skewness, as measured by the L-scale and the L-skewness respectively, in Proposition 3.5.

PROPOSITION 3.5 *(i) The L-scale of the equilibrium distribution is*

$$\lambda_{2,F} = \frac{(n-m)\mu}{n+1}. \tag{3.26}$$

(ii) The L-skewness of the equilibrium distribution is

$$\tau_{3,F} = \frac{n-2m}{n+2}. \tag{3.27}$$

(iii) The L-scale and the L-skewness of F are both strictly increasing in n and strictly decreasing in m.

(iv) The equilibrium PDF is symmetric about its mean if $m/n = 1/2$, right-skewed if $m/n < 1/2$, and left-skewed if $m/n > 1/2$.

Proof: See Appendix. □

Proposition 3.5 shows that the equilibrium distribution is right-skewed when the contest is selective, left-skewed when the contest is inclusive, and lacks skew when exactly one half of contestants receive a prize. This result can be interestingly contrasted with Gaba, Tsetlin, and Winkler (2004) who find that when contestants are only allowed to choose among symmetric performance distributions, each contestant maximizes performance variance when the contest is selective, minimizes performance variance when the contest is inclusive, and is indifferent between all levels of variance when one half of them receive a prize. Their result is driven by the symmetry assumption that forces contestants with skewness preference to exhibit variance preference.

Proposition 3.5 also implies that increasing contest selectivity increases the dispersion of the equilibrium distribution, which suggests that contestants take riskier strategies when contest selectivity increases. In the next proposition, we further confirm this point by showing that an increase in selectivity will induce a *mean-preserving increase in risk* (MPIR) of contestant performance. MPIR is a comparative risk concept that is more robust than comparisons of riskiness based on comparisons of a measure of dispersion (such as variance or L-scale). The notion of MPIR is the equal-means case of *second-order stochastic dominance* over pairs of probability distributions (see Machina and Pratt 1997). Particularly, $G(\cdot)$ differs from $F(\cdot)$ by an MPIR if and only if $G(\cdot)$ is second-order stochastically dominated by $F(\cdot)$ and has the same mean as $F(\cdot)$.

Proposition 3.6 *Increasing selectivity of a contest either by increasing the number of contestants or by decreasing the number of prizes induces a mean-preserving increase in risk of contestant performance.*

Proof: See Appendix. □

Contrasted with the findings from Gaba, Tsetlin, and Winkler (2004) that increasing selectivity either has no or an extreme effect on contestants' risk-taking behavior when contestants are restricted to symmetric distributions, Proposition 3.6 implies that the change in risk-taking is incremental in the change in selectivity when the symmetry restriction is removed.

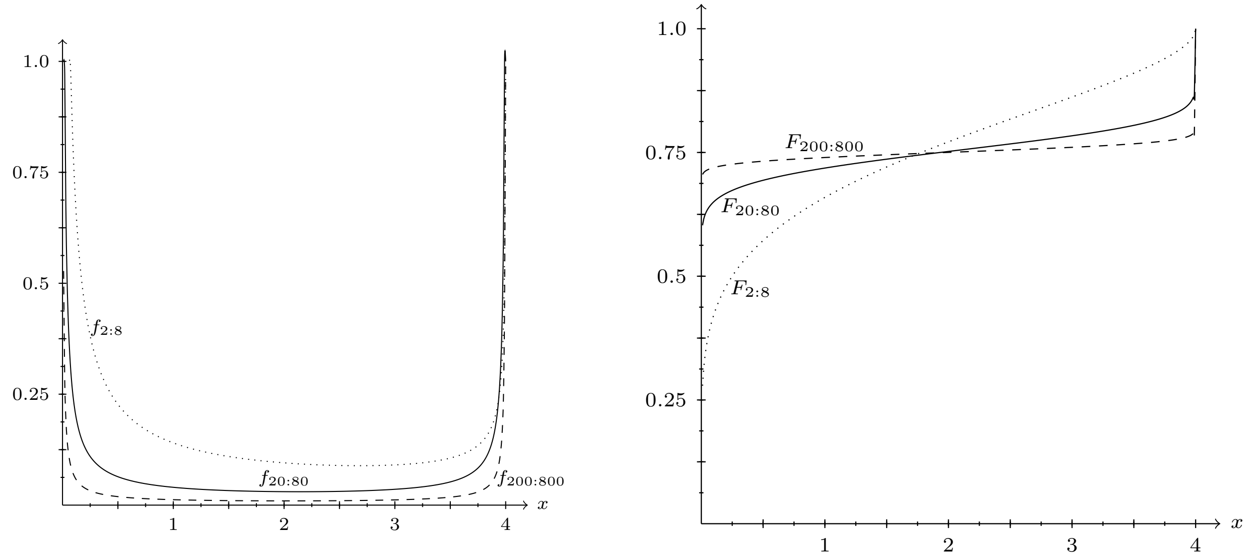
3.3.4 Contest Size

Contest size increases when n increases in proportion to m, with capacity fixed. In contrast to increasing selectivity, increasing contest size does not affect the support of the equilibrium distribution. However, increasing contest size affects the shape of the distribution. Using simple calculations based on the properties of the complementary beta distribution we show, in Proposition 3.7, that increasing contest size increases both the absolute value of skewness and the dispersion of equilibrium distributions.

Proposition 3.7 *When both n and m are multiplied by the same constant ρ, where $\rho > 1$ and ρn and ρm are both integers, there is no change in either the support of the equilibrium distribution or the direction of skewness, but both the L-scale and the absolute value of the L-skewness are increased.*

Proof: By Proposition 3.3, the support of the distribution is $[0, \frac{n\mu}{m}]$ and by Proposition 3.5, the sign of skewness is given by the sign of $(\frac{n}{m} - 2)$. Both values remain the same when n and m are both multiplied by the same constant. The rest of the proof is straightforward by comparing $\lambda_{2,F}(n,m)$ with $\lambda_{2,F}(\rho n, \rho m)$ and by comparing $\tau_{3,F}(n,m)$ with $\tau_{3,F}(\rho n, \rho m)$, where both $\lambda_{2,F}$ and $\tau_{3,F}$ are given by Proposition 3.5. We omit the computation. $\square$

The effects identified in Proposition 3.7 are depicted in Figure 3.4. In all the three cases depicted in Figure 3.4, one fourth of contestants in the contest win a prize. The PDFs and CDFs plotted vary by number of contestants, which varies in multiples of ten between eight and eight hundred.



Panel A. PDF Panel B. CDF

Figure 3.4: *Effect of increasing contest size on the equilibrium distribution.* The figure plots the equilibrium distributions for three contest sizes; $f_{i:j}$ ($F_{i:j}$) represents the equilibrium PDF (CDF) when i winners are selected out of j contestants. In each of the cases graphed, the winner proportion is fixed at 1/4.

Figure 3.4 illustrates that even large (ten-fold) increases in participation can have a fairly limited effect on the equilibrium performance distributions. All of the equilibrium distributions are confined to the same compact support. Besides, Figure 3.4 seems to indicate convergence to a limiting distribution as contest size increases without bound. This conjecture is indeed correct, as the next proposition shows.

Proposition 3.8 *For any sequence of natural numbers $(n_j, m_j)_j$ where $n_j \to \infty$ and for all j, $m_j = \rho n_j$, where $0 < \rho < 1$. Let $(F_j)_j$ be a corresponding sequence of equilibrium distributions when n_j is the number of contestants, m_j is the number of prize winners, and μ is the capacity of the contestants. The sequence of distributions $(F_j)_j$ converges* *weakly to the limiting distribution, F_∞, defined by*

$$F_\infty(x) = \begin{cases} 1-\rho & x < \mu/\rho \\ 1 & x \geq \mu/\rho \end{cases}. \tag{3.28}$$

Proof: See Appendix. □

The limiting distribution, F_∞, is Bernoulli, placing all its weight on the extreme points of the common support for the sequence of equilibrium distribution functions. The logic behind Proposition 3.8 is fairly straightforward: Holding the winner proportion and capacity constant while increasing the number of contestants makes the performance level required to win a prize more predictable. To counter this effect, the equilibrium distribution must become more unpredictable. Because contest size has no effect on the equilibrium range of performance levels, reduced predictability can only be produced by moving probability mass toward the extreme points of the support. In the limit, all weight is placed on these extreme points. Contrasted with the finding from Gaba, Tsetlin, and Winkler (2004) that contestants, who are restricted to symmetric distributions, play Bernoulli distributions only when the contest is selective, our result here shows that, after we remove the symmetry assumption, Bernoulli distribution is the limiting distribution no matter whether the contest is selective or inclusive.

3.4 Contests with Capacity Uncertainty

3.4.1 Probability of Winning

Many real-world contests are used to screen contestants and to promote winners who are supposed to be stronger than losers. These contests involve ex ante indistinguishable contestants with potential variations in their capacities. For these contests, it arises naturally the question "how efficient is the contest in selecting strong contestants?" This is the question concerned by employers who want to promote strong workers to the higher position in the hierarchy as the outputs there can be more sensitive to workers' capacities (see Meyer 1991). Besides, when employers recruit new workers by using contests, they want to select the most able contestants so as to maximize the future performance of new workers.

To address this question, we introduce capacity uncertainty into the model by assuming that each contestant has probability θ of being a strong (S) type, with capacity equal to μ_S, and probability $(1-\theta)$ of being a weak (W) type, with capacity equal to μ_W, where $0 < \mu_W < \mu_S$. A contestant's type is assumed to be the contestant's private information and is independent of the types of other contestants. Except for this uncertainty with respect to contestants' capacities, the contest game remains the same as the one defined at the beginning of Section 3.3.

In a symmetric Nash equilibrium, the probability of winning function, P, faced by all the contestants is the same. Each contestant must play a best reply to this function conditioned on his type. Hence, by symmetry, the set of best reply distributions is the same for all contestants conditioned on contestant type. Thus, without any possibility of confusion we will discuss the best reply and optimal strategy for type $t \in \{S, W\}$, recognizing that we are in fact referring to the best reply for any contestant whose type is t. Our first result generalizes Lemma 3.2, which shows that in any symmetric Nash equilibrium, no matter how many types a contestant could be, P is always continuous and intersects the origin.

LEMMA 3.4 *In a symmetric Nash equilibrium, the probability of winning function, P, is continuous and $P(0) = 0$.*

Proof: Suppose that P is not continuous. In which case, it must have a point of discontinuity, say $x^o \geq 0$. This implies that at least one type's equilibrium distribution must have a point of discontinuity at x^o, which further implies that this type places positive mass at x^o, so x^o must be a best reply for this type. However, by adopting the same argument as that in the proof of Lemma 3.2, we can show that this type is always better off transferring mass from x^o to $x^o + \epsilon$, for ϵ sufficiently small. The contradiction implies that P must be continuous. As no one places point mass, it is evident that $P(0) = 0$. $\square$

Thus, P satisfies the conditions for the maximization problem of Section 3.2. By (3.7), P satisfies the following condition: There exist nonnegative scalars $\alpha_t,\ \beta_t$ such that

$$P(x) \leq (\alpha_t + \beta_t x) \quad \forall\, x \geq 0 \tag{3.29}$$

$$\mathrm{Supp}_t \subset \{x \geq 0 : P(x) = (\alpha_t + \beta_t\, x)\}, \tag{3.30}$$

where Supp_t represents the support of F_t, the distribution selected by type t. Let ψ represent the concave lower envelope of the two upper support lines, $\{\alpha_t + \beta_t\, x\}_{t=S,W}$ associated with the two types, i.e.,

$$\psi(x) = \min[\alpha_S + \beta_S\, x, \alpha_W + \beta_W\, x]. \tag{3.31}$$

Next, note that the definition of the concave envelope, ψ, implies that

$$\forall\, t \in \{S, W\}, \quad \alpha_t + \beta_t\, x \geq \psi(x) \geq P(x), \tag{3.32}$$

i.e., the concave lower envelope of the two support lines lies between the support lines and P. Equations (3.30) and (3.32) imply that the supports of the two types' equilibrium distributions lie on the concave lower envelope. In fact, using Lemma 3.1 we can show that the probability of winning function, P, traces out the concave lower envelope generated by the two types' support lines until P reaches 1. The intuition behind the proof is that P can only grow at points in the supports of S and W's distributions. Because the equilibrium distributions' supports rest on points, x, at which the envelope meets P, to stay on the envelope, P can never increase at a rate in excess of the envelope's rate of increase. As soon as P breaks contact with the envelope, it must stay below the envelope because, by (3.30), it cannot ever increase again. Admittedly, this argument is a bit loose, but it captures the essence of the formal proof.

PROPOSITION 3.9 *There exist nonnegative constants, $\alpha_S, \alpha_W, \beta_S, \beta_W$ and $\hat{x} > 0$ such that if P represents the probability of winning function for a contestant in a symmetric Nash equilibrium,*

$$P(x) = \begin{cases} \min[\alpha_S + \beta_S\, x, \alpha_W + \beta_W\, x] & \text{if } x \leq \hat{x} \\ 1 & \text{if } x > \hat{x}, \end{cases} \tag{3.33}$$

where $\hat{x}$ is defined by

$$\min[\alpha_S + \beta_S\, \hat{x}, \alpha_W + \beta_W\, \hat{x}] = 1. \tag{3.34}$$

Proof: See Appendix. □

In essence, the probability of winning function, P, is the distribution a contestant plays against. As P is weakly concave, according to our finding in Section 3.2 that playing safe is a best reply to any weakly concave continuous distribution, a type-t contestant's probability of winning can be evaluated at his capacity, μ_t. Thus, the probability of winning for a type-t contestant equals $P(\mu_t)$ and the weak concavity of P implies the weak concavity of the value of capacity. We thus obtain the following corollary.

COROLLARY 3.1 *The value of capacity is weakly concave.*

Although we assumed two possible types of contestants for analytical convenience, Corollary 3.1 holds true for any number of possible types, because no matter how many types there are, the probability of winning function, P, always traces out the concave lower envelope of the support lines of all types until P reaches 1. Thus P is always weakly concave, implying the weak concavity of the value of capacity. Corollary 3.1 only requires that P is the same for all contestants, which is guaranteed when contestants are homogeneous ex ante.

Since the support of P equals the union of the supports of two types' equilibrium distributions. The following corollary is evident from Proposition 3.9.

COROLLARY 3.2 *The union of the supports of the equilibrium distributions of the two types equals* $[0, \hat{x}]$.

Next, in Lemma 3.5 we provide a characterization of the constants in the maximization problems that define the support lines. These constants are generated by the dual problem to the maximization problem given by expression (3.2) for each type. Our results show that the support line of a weak contestant always intersects the origin and that the slope of the W-support line is always weakly greater than the slope of the S-support line.

LEMMA 3.5 *The Lagrange multipliers from the dual problems satisfy the following constraints*

$$0 = \alpha_W \leq \alpha_S \tag{3.35}$$

$$0 < \beta_S \leq \beta_W \tag{3.36}$$

$$\mathrm{sgn}[\alpha_W - \alpha_S] = -\mathrm{sgn}[\beta_W - \beta_S]. \tag{3.37}$$

Proof: See Appendix. $\square$

Corollary 3.2 and Lemma 3.5 allow us to provide a complete characterization of the supports of the equilibrium distributions of the two types. This result is provided in Lemma 3.6.

LEMMA 3.6 *In any symmetric Nash equilibrium, there exists* $\hat{x}$ *such that the supports of the weak and strong types' distributions satisfy*

$$\begin{aligned} Supp_S &\subset \{x : x \leq \hat{x} \text{ and } \alpha_S + \beta_S\, x \leq \alpha_W + \beta_W\, x\} \\ Supp_W &\subset \{x : x \leq \hat{x} \text{ and } \alpha_S + \beta_S\, x \geq \alpha_W + \beta_W\, x\} \end{aligned} \tag{3.38}$$

where the constants, α_S, β_S, α_W, and β_W, satisfy the conditions in Lemma 3.5.

Proof: The result follows by noting that all points above the W-support line lie above P and thus cannot be best responses for S. The same argument shows that all points above the S-support line cannot be best responses for W. Thus the right-hand side of the two equations must enclose the support of the respective distributions. $\square$

3.4.2 Equilibrium Configuration

Lemmas 3.5 and 3.6 suggest that in equilibrium there are only two configuration candidates for the probability of winning function, P, which are illustrated in Figures 3.5 and 3.6. Figure 3.5 illustrates the case where $\alpha_W < \alpha_S$ and $\beta_W > \beta_S$. In this case, the support of the W's distribution lies strictly below that of the S's. In essence, a weak contestant concedes victory to strong contestants and wins only when matched against a sufficient number of weak contestants. We call equilibria with this configuration, "concession equilibria." Figure 3.6 illustrates the case where $\alpha_W = \alpha_S$ and $\beta_W = \beta_S$. In Figure 3.6, the support union of the two types' distributions equals $[0, \hat{x}]$, but each type's distribution support cannot be identified except in the special case where the upper bound of the W's distribution support coincides with the lower bound of the S's; in all the other cases, the upper bound of the W's distribution support lies strictly above the lower bound of the S's, implying that the performance distribution of a weak contestant challenges strong contestants and we thus call equilibria with this configuration, "challenge equilibria." As a weak contestant still concedes victory to the strong in the special case we just mentioned, to avoid any confusion, this special case is categorized into concession configuration.

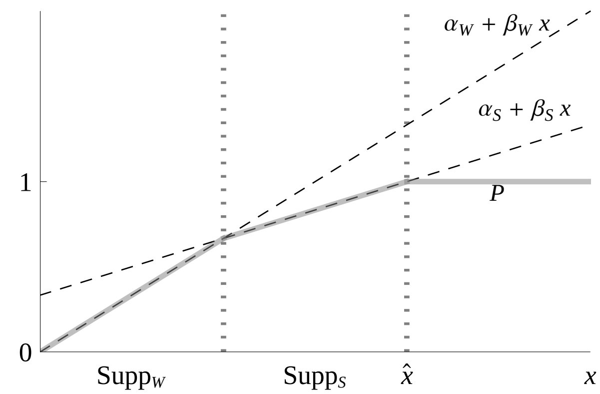


Figure 3.5: The probability of winning function, P, in concession equilibria

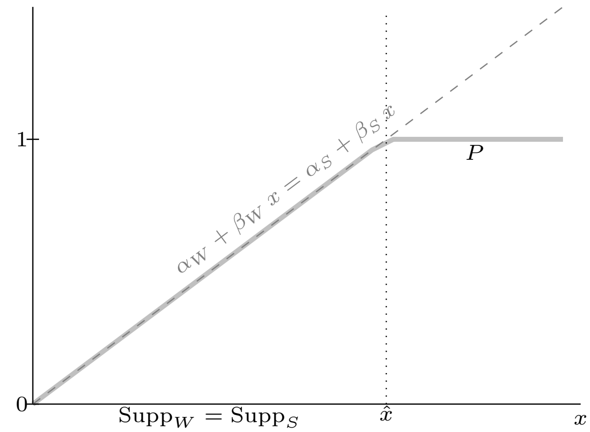


Figure 3.6: The probability of winning function, P, in challenge equilibria

The probability of winning function, P, for a contestant is determined by the equilibrium distributions of his competitors. Before we embark on systematic investigation of these equilibrium distributions, it will be useful to illustrate our result by a simple example. In the example, there are two contestants and one prize, i.e., $n = 2$ and $m = 1$. Ex ante, each contestant is equally likely to be strong or weak, i.e., $\theta = 1/2$. In this case, the probability that a given contestant wins with a realized performance level, x, in a symmetric equilibrium equals the probability that x is no less than the realized performance level of the other contestant. Since the other contestant is equally likely to be weak or strong, P is given by

$$P(x) = \frac{F_S(x) + F_W(x)}{2}. \tag{3.39}$$

In challenge equilibria, P is the CDF of a uniform distribution. Thus (3.39) requires that the average of distributions chosen by the two types equals a uniform distribution. This case is illustrated with the capacity of the weak type, μ_W, equal to 1 and the capacity of the strong type, μ_S, equal to 2 by Figure 3.7. Because the average capacity of the two types is 3/2 and because P is a uniform CDF,

$$\frac{x}{3} = \frac{F_S(x) + F_W(x)}{2} \quad x \in [0, 3]. \tag{3.40}$$

Note that the distributions, F_W and F_S, chosen by the two contestant types are not unique. All that is required for F_W and F_S to be equilibrium distributions for weak and strong types respectively is the satisfaction of equation (3.40). Figure 3.7 presents a particular choice of distribution functions satisfying (3.40).[12] Because P is a uniform distribution, all distributions with the same mean whose support is enclosed in P's produce the same payoff when used against P. Thus, we can evaluate the probability of winning of the weak and of the strong by evaluating P at μ_W and μ_S respectively. This yields a probability of winning equal to 1/3 for the weak type and 2/3 for the strong type. Note that the odds that the strong type will win, 2 : 1, equal the ratio of the strong type's capacity to the weak type's.

It is not always possible to satisfy the conditions for a challenge equilibrium. For example, consider the case where $\mu_W = 1$ and $\mu_S = 5$. Given these parameters, a challenge equilibrium would require that

$$\frac{x}{6} = \frac{F_S(x) + F_W(x)}{2} \quad x \in [0, 6], \tag{3.41}$$

which implies that

$$\frac{x}{6} \geq \frac{F_W(x)}{2}. \tag{3.42}$$

Integrating both sides of (3.42) with respect to $F_W(x)$ over the support of F_W yields

$$\int \frac{x}{6} dF_W(x) \geq \int \frac{F_W(x)}{2} dF_W(x). \tag{3.43}$$

[12]In Figure 3.7, the distributions chosen by weak and strong contestants are as follows: the weak type plays the uniform distribution on $[0, 3/2]$ with probability 5/6 and plays the uniform distribution on $[3/2, 3]$ with probability 1/6; the strong type plays the uniform distribution on $[0, 3/2]$ with probability 1/6 and plays the uniform distribution on $[3/2, 3]$ with probability 5/6.

Because $\mu_W = 1$, the left-hand side of (3.43) evaluates to 1/6. However, the right-hand side of (3.43) equals 1/4. Thus, (3.43) cannot hold, so there exists no challenge equilibrium in this case. In essence, when the weak type's capacity is too small, weak contestants are unable to implement strategies that contribute to producing the uniform probability of winning distribution required for a challenge equilibrium. In this case only "concession equilibria" exist. A concession equilibrium is illustrated in Figure 3.8. In this symmetric equilibrium, weak contestants use a uniform distribution over [0, 2] while strong contestants use a uniform distribution over [2, 8]. A weak contestant can win only when the other contestant is also weak, which occurs one half of the time. Moreover, if both contestants are weak, both are using the same distribution and thus have an equal probability of winning. Hence, a weak contestant's probability of winning is $(1/2) \times (1/2) = 1/4$ and a strong contestant's probability is 3/4.

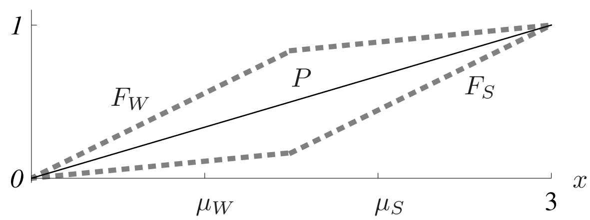


Figure 3.7: An example of a challenge equilibrium. In the example, $\mu_W = 1$, $\mu_S = 2$, the number of contestants, n, equals 2, and the number of winners, m equals 1. The probability a contestant is strong, θ, equals 1/2.

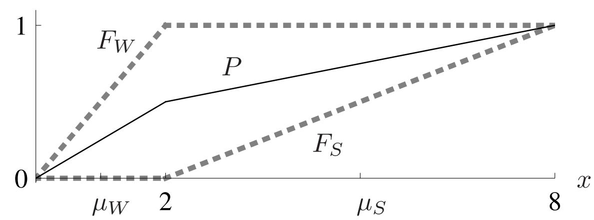


Figure 3.8: An example of a concession equilibrium. In the example, $\mu_W = 1$, $\mu_S = 5$, the number of contestants, n, equals 2, and the number of prizes, m equals 1. The probability a contestant is strong, θ, equals 1/2.

Now we begin to study the general case with $1 \leq m \leq (n-1)$, $0 < \theta < 1$, and $0 < \mu_W < \mu_S$. Recall that there are only two candidate equilibrium configurations: concession and challenge. We denote the concession configuration by C and denote by p_t^C the type-t contestant's expected probability of winning under the concession configuration, for $t \in \{S, W\}$. Because, under the concession configuration, weak contestants concede victory to strong contestants, the expressions of p_S^C and p_W^C can be worked out by referring to the prize-allocation rule that for any random draw of contestant types, strong contestants always have priority over weak ones to win and contestants of the same type have the same chance of winning. It is thus noticeable that p_S^C and p_W^C are uniquely determined by n, m, and θ. As we can analyze selection efficiency, which is the focus of this section, without presenting p_S^C and p_W^C, we omit writing down their expressions here.

We denote the challenge configuration by G and denote by p_t^G the type-t contestant's expected probability of winning under the challenge configuration, for $t \in \{S, W\}$. As under the challenge configuration, both types' upper support lines overlap and intersect the origin, $p_W^G : p_S^G = \mu_W : \mu_S$. By symmetry, ex ante each player has a chance of winning equal to m/n, so $\theta p_S^G + (1-\theta)p_W^G = m/n$. Hence,

$$\begin{aligned} p_S^G &= \frac{m\mu_S}{n\left[\theta\mu_S + (1-\theta)\mu_W\right]}; \\ p_W^G &= \frac{m\mu_W}{n\left[\theta\mu_S + (1-\theta)\mu_W\right]}. \end{aligned} \tag{3.44}$$

The next lemma shows that, which configuration is supported to be the equilibrium configuration for a given parameterization of the model is determined by which configuration favors the weak type.

Lemma 3.7 *If p_W^C and p_W^G represent the weak type's probability of winning under the concession and challenge configurations respectively, concession equilibria exist if and only if $p_W^C \geq p_W^G$; challenge equilibria exist if and only if $p_W^C < p_W^G$.*

Proof: See Appendix. □

The intuition of Lemma 3.7 is as follows: enabling contestants to choose performance distribution offers weak contestants the possibility of outperforming strong ones. However, challenging the strong may not always be optimal for a given weak contestant, since to challenge, this weak contestant must prolong the right tail of his performance distribution, which, through the capacity constraint, increases the probability of low performance, thus reducing his chance of winning when competing against weak opponents.

3.4.3 Selection Efficiency

Lemma 3.7 helps us examine selection efficiency of contests. Before we do this, we need to settle on a definition of selection efficiency. Selection efficiency of a mechanism is a characteristic of the mechanism but not the quality of contestants per se. Thus, when all contestants happen to be weak, even the most efficient mechanism will not be able to ensure that the winner is strong. A selection mechanism is efficient to the extent that it gives stronger contestants priority in winning. The most efficient selection mechanism is a mechanism under which strong contestants have absolute priority — i.e., the probability that a weak contestant wins a prize given that not all strong contestants win a prize is 0. For a given contest, *maximum selection efficiency* is the probability that a selected contestant will be strong under the most efficient selection mechanism. So, for example, if there are two contestants and one prize, and the probability that a given contestant is strong equals 1/2, then the most efficient mechanism will select a strong contestant whenever at least one of the two contestants is strong. Since 3/4 of the time at least one contestant is strong, maximum selection efficiency equals 3/4. We denote maximum selection efficiency by Π^*. We compare maximum selection efficiency with *actual selection efficiency*, denoted by Π. Actual selection efficiency is the equilibrium probability of a selected contestant being strong in a symmetric Nash equilibrium. The difference between Π^* and Π is called *selection efficiency loss*, denoted by $\Delta\Pi$, where $\Delta\Pi \equiv \Pi^* - \Pi$. A contest is said to be *efficient* if and only if $\Delta\Pi = 0$. While actual selection efficiency measures the quality of prize winners, selection efficiency loss measures the reduction in winner quality caused by a given contest mechanism. A contest mechanism owns poor selection properties if it produces large selection efficiency loss.[13]

In concession equilibria, strong contestants have priority over weak ones to win, so $\Pi = \Pi^*$. In challenge equilibria, each contestant's probability of winning is proportional to his capacity, so by Bayes' Rule, $\Pi = \frac{\theta r}{(1-\theta)+\theta r}$, where $r = \mu_S/\mu_W$, the ratio of the strong type's capacity to the weak type's. We interpret r as a measure of strength asymmetry. By Lemma 3.7, between concession and challenge configurations, the equilibrium configuration is always the one that favors the weak type, so actual selection efficiency, Π, equals the minimum of Π^* and $\frac{\theta r}{(1-\theta)+\theta r}$. The next proposition summarizes these results.

Proposition 3.10 *Maximum selection efficiency,* Π^**, actual selection efficiency,* Π*, and selection efficiency loss,* $\Delta\Pi$*, are given as follows:*

$$\Pi^* = \sum_{i=0}^{n} \binom{n}{i} \min\left\{\frac{i}{m}, 1\right\} \theta^i (1-\theta)^{n-i}. \tag{3.45}$$

$$\Pi = \min\left\{\frac{\theta r}{(1-\theta)+\theta r}, \Pi^*\right\}. \tag{3.46}$$

$$\Delta\Pi = \max\left\{\Pi^* - \frac{\theta r}{(1-\theta)+\theta r}, 0\right\}. \tag{3.47}$$

Proof: See Appendix. □

Using Proposition 3.10, we examine comparative statics on actual selection efficiency and selection efficiency loss. By equation (3.45), Π^* does not depend on μ_S or μ_W. Note that a change in either μ_S or μ_W can only affect actual selection efficiency through the change in strength asymmetry, r. Hence, instead of presenting comparative static results with respect to μ_S and μ_W, we present those with respect to r.

Corollary 3.3 *The comparative static results on* Π *and* $\Delta\Pi$ *are given as follows:*

(i) For fixed θ*,* n *and* m*, there exists a threshold level of strength asymmetry,* r^**, such*

[13] Our approach of using two metrics, actual selection efficiency and selection efficiency loss, contrasts with the previous literature including Hvide and Kristiansen (2003) and Ryvkin and Ortmann (2008) which only uses actual selection efficiency as the metric of efficiency.

that Π *is strictly increasing in* r *when* $r < r^*$ *and is constant in* r *for* $r \geq r^*$*, whereas* $\Delta\Pi$ *is positive and strictly decreasing in* r *when* $r < r^*$ *and equals zero when* $r \geq r^*$*.*

(ii) For fixed r*,* n*, and* m*,* Π *is strictly increasing in the probability of a contestant being strong,* θ*, whereas* $\Delta\Pi$ *is not monotonic in* θ*. There exists* $\hat{\theta} \in (0,1)$ *such that* $\Delta\Pi$ *is maximized at* $\hat{\theta}$*.*

(iii) For fixed θ*,* r*, and* m*, there exists a threshold number of contestants,* n^**, such that* Π *is strictly increasing in* n *when* $n \leq n^*$ *and is constant in* n *when* $n > n^*$*, whereas* $\Delta\Pi$ *equals zero when* $n \leq n^*$ *and is positive and strictly increasing in* n *when* $n > n^*$*.*

(iv) For fixed θ*,* r*, and* n*, if*

$$r \geq \frac{1-(1-\theta)^n}{\theta(1-\theta)^{n-1}}, \tag{3.48}$$

Π *is strictly decreasing in the number of prizes,* m*, and* $\Delta\Pi$ *equals 0. If* (3.48) *is violated, there exists a threshold number of prizes,* m^**, such that* Π *is constant in* m *for* $m < m^*$ *and is strictly decreasing in* m *for* $m \geq m^*$*, whereas* $\Delta\Pi$ *is positive and strictly decreasing in* m *when* $m < m^*$ *and equals zero when* $m \geq m^*$*.*

(v) Holding θ *and* r *fixed while increasing the scale of the contest by multiplying both the number of contestants,* n*, and the number of prizes,* m*, by a common integer factor,* k*, weakly increases both* Π *and* $\Delta\Pi$*.*

Proof: See Appendix. $\square$

Intuitively, a contest is efficient when weak contestants concede victory to strong contestants. Weak contestants are more likely to concede under the following situations: (i) an increase in the degree of strength asymmetry, measured by r, (ii) a decrease in the likelihood that a given opponent is strong, θ, (iii) a decrease in the number of contestants, n, or (iv) an increase in the number of prizes, m. In the opposite situations, weak contestants are more likely to challenge the strong, generating selection efficiency loss.[14]

[14]Selection efficiency loss, $\Delta\Pi$, is monotonic in r, n, and m, but nonmonotonic in θ. Increasing θ

To see how a change in contest parameter affects actual selection efficiency, we need to first look at its effect on maximum selection efficiency. The effect is purely statistical. As we measure selection efficiency by winners' types rather than by winners' absolute strength, a change in one type's absolute strength, μ_S and μ_W, or relative strength, r, does not affect maximum selection efficiency. An increase in the number of contestants, n, a decrease in the number of prizes, m, and an increase in each contestant's probability of being strong, θ, all improve maximum selection efficiency, because all of them statistically improve the expected quality of the top m contestants. Increasing the scale of a contest works in a similar but slightly more subtle fashion: when there are few contestants, then the realized proportion of strong contestants is more likely to deviate from its expected value, θ. If more strong contestants are drawn than prizes, the excess of strong contestants has no positive effect on the quality of prize winners but if less are drawn it has a negative effect. Thus, random variation in the quality of the contestant pool lowers maximum selection efficiency. By the law of large numbers, scaling up the contest reduces random variation in the realized fraction of strong contestants and thus increases maximum selection efficiency.

Actual selection efficiency is the difference between maximum selection efficiency and selection efficiency loss. It is quite often that a change in parameter, such as a change in n or m, affects maximum selection efficiency, through statistical effect, and affects selection efficiency loss, through strategic effect, in the same direction. However, our results imply that whenever this happens, the statistical effect is always the weakly dominant effect. Thus, the effect of a change in parameter on actual selection efficiency is always (weakly) monotonic. This result can be contrasted with Hvide and Kristiansen (2003), who find that an increase in n or θ can sometimes decrease actual selection efficiency. Their result is largely driven by the assumption that contestants can only choose between a constant

increases a weak contestant's benefit from challenging the strong. Once this benefit exceeds a threshold, only challenge equilibria can be supported. This effect raises $\Delta\Pi$ above zero. However, as θ continues to increase toward 1, the probability that a contestant is weak decreases to 0, so $\Delta\Pi$ approaches zero again.

and a Bernoulli-distributed random variable. This assumption prevents strong contestants from using the win-small/lose-big strategy to better accommodate to the challenge brought by weak contestants and thus amplifies the negative strategic effect on actual selection efficiency when n or θ increases.

3.5 Selection Efficiency of Modified Contest Mechanisms

In Section 3.4, we studied selection efficiency of a simple contest mechanism in which contestants compete against each other for a certain number of prizes with no restrictions imposed on admissible performance distributions apart from non-negativity and capacity. In this section, we treat this simple contest mechanism as the benchmark and study whether a principal who seeks to maximize selection efficiency could do better with some small modifications of the benchmark mechanism. As Corollary 3.3 in Section 3.4 has already presented comparative statics on selection efficiency with respect to the number of contestants, the number of prizes, the quality of contestants, and the strength asymmetry, to avoid repetition of analysis, in this section, we assume that all these contest parameters are fixed for the principal. However, even if the principal cannot change these parameters, the principal still has various ways to modify the benchmark mechanism and it is impossible for us to enumerate all of them. Thus, we restrict our focus to three alternative contest arrangements that are close to the benchmark mechanism and commonly observed in practice. We aim to find out whether these alternative arrangements can help improve selection efficiency.

As we assume that the parameters studied in Corollary 3.3 are fixed, maximum selection efficiency is fixed, in which case the two metrics, actual selection efficiency and selection efficiency loss, used for evaluating selection efficiency are equivalent. Thus, in what follows, we base our analysis of selection efficiency on the analysis of actual selection efficiency.

3.5.1 Scoring Caps

Suppose contestants' performance is capped by a certain level, $\bar{x}$, so that contestants are restricted to distributions satisfying $F[\bar{x}] = 1$. The use of a scoring cap imposes an upper bound on contestants' performance levels. There are many real-life contests with a scoring cap. For example, in examinations and many sports games, such as gymnastics and shooting, contestant performance levels are bounded by full scores. A principal can change this upper bound by changing the difficulty of reaching a full score. A scoring cap is fairly easy for the principal to enforce. The principal only needs to specify that all performance levels greater than or equal to $\bar{x}$ will be treated the same for the purpose of determining contest winners. Under this specification, contestants have no incentive to assign positive probability to performance levels exceeding $\bar{x}$. However, imposing a scoring cap does require credible commitment by the principal and communication of the rule to contestants ex ante.

When $\bar{x}$ is strictly less than the strong type's capacity, μ_S, strong contestants will not fully utilize their capacity; neither will weak contestants when $\bar{x}$ is strictly less than the weak type's capacity, μ_W. However, it is obvious that the imposition of $\bar{x} \in (0, \mu_W)$ handicaps strong contestants as it makes strong contestants only able to utilize the same amount of capacity as that of weak ones. Hence the imposition of $\bar{x} \in (0, \mu_S)$ would (weakly) hurt selection efficiency.[15] In addition, imposing $\bar{x} < \mu_S$ would reduce the expected total performance. Thus, a principal who cares about selection efficiency or total performance would never set $\bar{x}$ below μ_S. Hence in what follows, we focus on the case where $\bar{x} \geq \mu_S$.

With the imposition of a scoring cap, $\bar{x}$, the argument in the proof of Lemma 3.4 can

[15] If with the imposition of $\bar{x} < \mu_S$, weak contestants still prefer conceding victory to the strong, selection efficiency remains the same.

be applied to every point on $[0, \bar{x})$ to show that in any symmetric Nash equilibrium no contestant places point mass on $[0, \bar{x})$. Thus, the probability of winning function, P, is continuous on $[0, \bar{x})$ and intersects the origin. However, the same argument cannot be applied to the scoring cap $\bar{x}$. Thus, there may be point mass on $\bar{x}$ and the continuity of P may break at $\bar{x}$. The reason that a point mass on $\bar{x}$ can be supported in equilibrium is that the scoring cap x prevents a contestant from topping $\bar{x}$ by an infinitesimal positive amount which vastly increases his chance of winning with a negligible impact on capacity constraint. We summarize our observation in the following lemma, whose proof can be referred to the proof of Lemma 3.4.

LEMMA 3.8 *With the imposition of a scoring cap, $\bar{x}$, in a symmetric Nash equilibrium, the probability of winning function, P, is continuous on $[0, \bar{x})$ and intersects the origin but may not be continuous at $\bar{x}$.*

Although a discontinuity can occur at $\bar{x}$, it is still the case that the support of contestants' performance distributions must lie on the lines of support for the dual problem. Thus, if without the scoring cap $\bar{x}$, a contestant places all weight over a support which contains $\bar{x}$ in equilibrium, then after the imposition of $\bar{x}$, this contestant will move all the weight placed above $\bar{x}$ as well as some weight placed below $\bar{x}$ to the point mass on $\bar{x}$. Since this transformation does not affect any type's upper support line, it does not affect any type's probability of winning. Therefore, we have the following result.

PROPOSITION 3.11 *The imposition of a scoring cap, weakly greater than the strong type's capacity, does not affect any type's probability of winning or selection efficiency.*

Proof: See Appendix. □

Although the use of a scoring cap $\bar{x} \geq \mu_S$ does not affect selection efficiency, it changes equilibrium distributions: any weight that is originally placed above $\bar{x}$ is now moved to $\bar{x}$ and to balance its effect on the mean, some weight that is originally placed below $\bar{x}$ is also moved to $\bar{x}$. If we lower $\bar{x}$, more and more weight will be placed on $\bar{x}$ rather than spread over the neighborhood of $\bar{x}$, resulting in a mean-preserving contraction of contestants' equilibrium distributions. Hence, a principal who is averse to performance riskiness of contestants prefers a tighter scoring cap whenever the scoring cap exceeds μ_S.[16]

PROPOSITION 3.12 *A principal who is risk averse in contestants' performance levels weakly prefers a tighter scoring cap, provided that the scoring cap is weakly greater than the capacity of strong contestants, μ_S, and the optimal scoring cap equals μ_S.*

Proof: See Appendix. $\square$

The use of a scoring cap can reduce contestants' performance riskiness with no "side effects" of bringing down selection efficiency as long as the scoring cap exceeds strong contestants' capacity. Thus, the use of a scoring cap is beneficial to a principal who is risk averse to contestants' performance levels, even if she also cares about selection efficiency. Note that a precise knowledge on μ_S is not required for the implementation of a scoring cap without hurting selection efficiency; if the principal knows that $\mu_S \in [a, b]$, she can simply set the scoring cap at b without any worries about selection efficiency.

3.5.2 Penalty Triggers

In the previous sections, contestants are not penalized for bad performance. In this section, we study whether selection efficiency can be improved by the use of penalty triggers: a contestant is penalized if his performance level falls strictly below a threshold, $\underline{x}$. We look at the case where $\underline{x} \leq \mu_W$ so that it is possible for weak contestants to perfectly avoid the penalty. We assume that the penalty is sufficiently large relative to a winner's prize so that no contestant has an incentive to place any weight below $\underline{x}$. Thus, the use of penalty triggers imposes a lower bound, $\underline{x}$, on performance levels chosen by contestants.

[16]See Diamond and Stiglitz (1974) and Machina and Pratt (1997) for the definition of mean-preserving contraction and a discussion of how it is related to risk-averse preferences.

Intuitively, when penalty triggers are used, we can think of each type-t contestant performance as the sum of two parts: a safe performance level equal to $\underline{x}$, submitted for the purpose of avoiding the penalty, and a random performance level whose expected value equals $(\mu_t - \underline{x})$ for all $t \in \{S, W\}$, submitted for the purpose of maximizing the chance of winning a prize. As every contestant spares $\underline{x}$ units of capacity to avoid the penalty, contestants are in fact competing by transforming the rest $(\mu_t - \underline{x})$ units of capacity into a distribution and the contest is essentially the same as the one studied in Section 3.4 except that now each type-t contestant's "manipulatable" capacity is $(\mu_t - \underline{x})$ instead of μ_t, for all $t \in \{S, W\}$. An increase in $\underline{x}$ increases $(\mu_S - \underline{x})/(\mu_W - \underline{x})$, the ratio of strong contestants' manipulatable capacity to the weak's, which effect is equivalent to an increase in the strength asymmetry, measured by r, in the original contest. Thus, the next proposition, which shows that the use of penalty triggers (weakly) improves selection efficiency, is straightforward from (i) in Corollary 3.3.

PROPOSITION 3.13 *Increasing the threshold, $\underline{x}$, that triggers the penalty strictly improves selection efficiency as long as $\frac{\mu_S - \underline{x}}{(\mu_W - \underline{x})} \leq r^*$, where r^* is defined in Corollary 3.3; otherwise, an increase in $\underline{x}$ has no effect on selection efficiency.*

Remark: When $\underline{x} = \mu_W$, treat $(\mu_S - \underline{x})/(\mu_W - \underline{x})$ as ∞.

Proposition 3.13 suggests that, if a principal wants to promote the most able employees to let them take over the higher hierarchy work, penalty triggers would be beneficial to selection efficiency. Proposition 3.13 thus explains why mixed incentive systems with both "carrots" and "sticks" are often found in practice (Grote 2002).[17]

Since the use of penalty triggers can improve selection efficiency, which is the major concern of the principal in our model, we only briefly talk about its impact on the riskiness of performance levels. We use an example below to show that the impact can be

[17]Other studies on prizes and penalties in contests include Gilpatric (2009), Akerlof and Holden (2010), and Moldovanu, Sela, and Shi (2010). But in these papers, the principal's goal is to induce contestants' effort rather than maximize selection efficiency. Another difference is that the penalties in these papers are rank-dependent, but in our model, they are performance-dependent.

ambiguous.

Suppose without penalty triggers, weak contestants make concession to the strong. Then implied by Proposition 3.13, after the penalty-triggering threshold, $\underline{x} > 0$, is imposed, weak contestants will still make concession to the strong. Figure 3.9 illustrates the probability of winning functions with and without $\underline{x} > 0$ in such a case, denoted by P' and P respectively. The upper bounds of the support of weak contestants' distributions with and without $\underline{x} > 0$ are denoted by $\check{x}'$ and $\check{x}$ respectively and those of strong contestants' are denoted by $\hat{x}'$ and $\hat{x}$ respectively. Figure 3.9 has three properties: (i) $P(\check{x}) = P'(\check{x}')$, (ii) $\check{x}' < \check{x}$, and (iii) $\hat{x}' > \hat{x}$. The reason for (i) is that both $P(\check{x})$ and $P(\check{x}')$ equal the probability of winning if a contestant concedes victory to strong contestants but has priority over weak contestants to win. Property (i) implies property (ii), since otherwise a weak contestant's distribution with $\underline{x} > 0$ imposed will first-order stochastically dominate that without $\underline{x}$, resulting in difference between the weak type's capacities with and without $\underline{x}$. Property (ii) further implies property (iii), since otherwise a strong contestant's distribution with $\underline{x} > 0$ imposed will be first-order stochastically dominated by that without $\underline{x}$, resulting in difference between the strong type's capacities with and without $\underline{x}$. Owning these properties, Figure 3.9 illustrates a case where imposing the penalty-triggering threshold $\underline{x} > 0$ induces a mean-preserving reduction in risk of weak contestants' performance but a mean-preserving increase in risk of strong contestants' performance. Since P' and P, which are convex combinations of strong and weak contestants' performance distributions with and without penalty triggers respectively, do not satisfy the single-crossing property, the aggregate effect on performance riskiness is ambiguous in this example.

In contrast to the use of scoring caps, whose effects are presented in Propositions 3.11 and 3.12, the use of penalty triggers increases selection efficiency but has an ambiguous effect on the riskiness of contestants' performance. Thus, penalty triggers play a different role from scoring caps. Scoring caps would be used to reduce the riskiness of contestants'

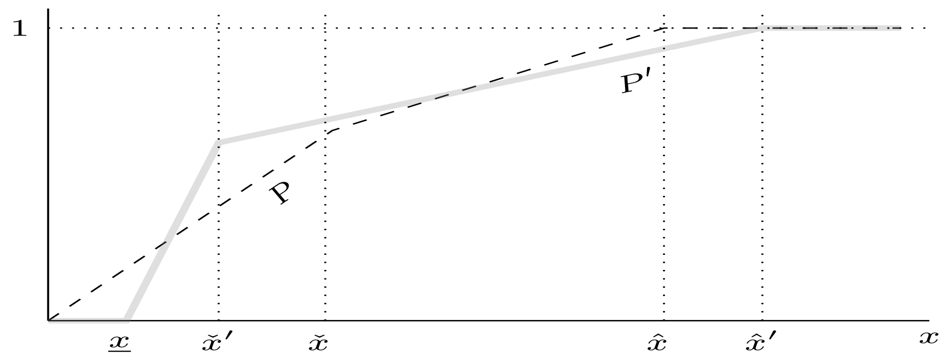


Figure 3.9: Concession equilibria with and without penalty triggers

performance while penalty triggers would be used to improve selection efficiency. Thus, scoring caps and penalty triggers are complements rather than substitutes. For this reason, it is not surprising that actual contest designs frequently use both.

3.5.3 Localizing Contests

In this section, we study the effect on selection efficiency of dividing the original grand contest involving n contestants and m prizes into multiple smaller local contests with contestants in the same local contest competing for prizes allocated to that local contest. An admissible division of local contests requires the number of contestants in each local contest summing up to n while the number of prizes summing up to m. The next proposition shows that localizing contests weakly lowers selection efficiency.

PROPOSITION 3.14 *Localizing a contest never improves selection efficiency. The effect is neutral if and only if every local contest has challenge equilibria.*

Proof: See Appendix. $\square$

Proposition 3.14 shows the beauty of "bigness": to maximize selection efficiency, it is weakly optimal to use a grand contest rather than localizing it. The intuition is easiest to understand if we look at the effect of the opposite operation – grouping local contests together. Grouping local contests acts in a similar way as scaling up those local contests, which by (v) in Corollary 3.3, weakly increases selection efficiency. When every local contest has challenge equilibria, implying that the competition in every local contest is so intense that weak contestants cannot afford conceding victory to the strong, grouping local contests into a bigger contest will never bring down the intensity of competition, so the bigger contest will also induce challenge equilibria. Note that in challenge equilibria, actual selection efficiency equals $\frac{\theta r}{(1-\theta)+\theta r}$, which is independent of contest size. Thus, selection efficiency remains the same by grouping local contests with challenge equilibria.

3.6 Conclusion

In this paper, we studied contests where, subject only to a capacity constraint on mean performance, contestants chose arbitrary performance distributions. In the case of symmetric capacity, we derived closed-form solutions for equilibrium performance distributions and analyzed the effect of contest structure on equilibrium behavior. We showed that equilibrium performance distributions are never unimodal (unless the contest produces only one winner or one loser) and are always right skewed when the contest is selective. We then extended the analysis to consider the case where contestants are unaware of each others' capacities. In this setting, we characterized equilibria and analyzed the effect of changing contest parameters on strategies, payoffs, and overall contest efficiency. We showed, contrary to the ruin-and-risk-taking intuition, that weaker contestants do not always gamble on high-risk strategies and that, when the capacities of weak and strong contestants are sufficiently different, the contest mechanism produces perfect selection efficiency. We then considered the effect of various modifications of the contest mechanism.

Given the simplicity of our underlying model structure, there is significant room to extend this analysis without losing tractability. One potential direction for extension is to endogenize capacity. Capacity could be endogenized through a two-stage model, with costly contest capacity creation in the first stage followed by capacity-constrained distribution choice in the second stage. The tension in such a model would be that increased capacity investments by stronger contestants increase selection but also dissipate resources. Another direction of extension would be to consider asymmetric contests where the contestants know the capacity of other contestants. This extension could represent contests between socially connected contestants with intimate knowledge of each others' abilities, e.g., insider contests for CEO succession. Finally, we could extend the analysis by explicitly modeling the preferences of the contest designer and examining how these preferences affect contest design parameters. This extension could address issues such as the dynamic consistency of the designer's ex ante preference for contest-based selection with the designer's ex post preference conditioned on observed contestant actions.

3.7 Appendix: Proofs

Proof of Lemma 3.1: Define

$$\mathcal{S} = \{x \in [0, c) : P(x) = \bar{P}(x)\} \quad \mathcal{S}^c = \{x \in [0, c) : P(x) < \bar{P}(x)\} \tag{3.49}$$

Then, $\mathcal{S}^c$ and $\mathcal{S}$ are disjoint sets and

$$[0, c) = \mathcal{S}^c \cup \mathcal{S}. \tag{3.50}$$

By assumption,

$$dP(\mathcal{S}^c) = 0. \tag{3.51}$$

For any measurable set $\mathcal{E}$,

$$\mathcal{E} = (\mathcal{E} \cap \mathcal{S}^c) \cup (\mathcal{E} \cap \mathcal{S}) \tag{3.52}$$

By (3.51), we see that

$$dP(\mathcal{E}) = dP(\mathcal{E} \cap \mathcal{S}) = d\bar{P}(\mathcal{E} \cap \mathcal{S}) \tag{3.53}$$

Thus,

$$dP(\mathcal{E}) = dP(\mathcal{E} \cap \mathcal{S}) = d\bar{P}(\mathcal{E} \cap \mathcal{S}) \leq d\bar{P}(\mathcal{E}) \tag{3.54}$$

By assumption (ii), there exists x' such that $P(x') < \bar{P}(x)$. Using the definition of P we see that this implies that

$$dP([0, x']) < d\bar{P}([0, x']). \tag{3.55}$$

For all $x > x'$ is must be the case that

$$\begin{aligned} P(x) &= dP([0, x']) + dP((x', x]) \\ \bar{P}(x) &= d\bar{P}([0, x']) + d\bar{P}((x', x]), \end{aligned} \tag{3.56}$$

and by (3.54)

$$dP((x', x]) \leq d\bar{P}((x', x]). \tag{3.57}$$

(3.55) and (3.57) imply that,

$$P(x') < \bar{P}(x') \;\&\; x > x' \Rightarrow P(x) < \bar{P}(x) \tag{3.58}$$

This implies by (3.13) that P is constant over (x', c). Thus, $\mathcal{S}$ is a closed interval of the form $[0, \hat{x}]$ and thus

$$P(x) = \int_0^{\min[x, \hat{x}]} d\bar{P}. \tag{3.59}$$

and p^* is defined by

$$p^* = dP[0, c). \tag{3.60}$$

□

Proof of Proposition 3.5: This result is a straightforward implication of the following result:

RESULT 1 *(Jones 2002) The L-scale of $CB(a, b)$ is*

$$\lambda_{2,CB} = \frac{a\,b}{(a+b)(a+b+1)}. \tag{3.61}$$

The L-skewness of $CB(a, b)$ is

$$\tau_{3,CB} = \frac{a-b}{a+b+2}. \tag{3.62}$$

Proof: See Jones (2002). □

Substituting $(n-m)$ for a and m for b in equations (3.61) and (3.62) gives $\lambda_{2,CB} = \frac{(n-m)m}{n(n+1)}$ and $\tau_{3,CB} = \frac{n-2m}{n+2}$. As $\frac{m}{n\mu}X \sim CB(n-m, m)$, X has the L-scale $\lambda_{2,F}$ equal to $\lambda_{2,CB}$ multiplied by $(n\mu)/m$, which gives equation (3.26), and X has the L-skewness $\tau_{3,F}$ equal to $\tau_{3,CB}$ as the L-skewness is scale invariant, which gives equation (3.27). Part (iii) then follows from examining the signs of the first orders of $\lambda_{2,F}$ and $\tau_{3,F}$ with respect to n and m and part (iv) from examining the sign of $\tau_{3,F}$. We omit the calculations. $\square$

Proof of Proposition 3.6: Let $F_{(n,m)}(\cdot)$ represent the equilibrium distribution where there are n contestants competing for m prizes. Suppose $n' > n$ and $m' < m$, so both contest (n', m) and contest (n, m') are more selective than contest (n, m). As in equilibrium, each contestant fully utilizes capacity, $F_{(n,m)}$, $F_{(n',m)}$, and $F_{(n,m')}$ have the same mean. Thus, to prove this proposition, all we need to show is that $F_{(n,m)}$ second-order stochastically dominates both $F_{(n',m)}$ and $F_{(n,m')}$. We do this by showing that (a) there is a unique quantile $\hat{q} \in (0,1)$ with the property that the horizontal distance between $F_{(n',m)}$ and $F_{(n,m)}$, which is $F^{-1}_{(n',m)}(q) - F^{-1}_{(n,m)}(q)$, is strictly negative if and only if $q < \hat{q}$, and, in a like fashion, (b) there is a unique quantile $\bar{q} \in (0,1)$ with the property that $F^{-1}_{(n,m')}(q) < F^{-1}_{(n,m)}(q)$ if and only if $q < \bar{q}$.

We first prove (a). By Proposition 3.3,

$$F^{-1}_{(n',m)}(q) - F^{-1}_{(n,m)}(q) = \frac{n'\mu}{m} \sum_{i=n'-m}^{n'-1} \binom{n'-1}{i} q^i (1-q)^{(n'-1)-i} - \frac{n\mu}{m} \sum_{i=n-m}^{n-1} \binom{n-1}{i} q^i (1-q)^{(n-1)-i}. \tag{3.63}$$

Differentiating equation (3.63) with respect to q gives

$$\begin{aligned} &\frac{d(F^{-1}_{(n',m)}(q) - F^{-1}_{(n,m)}(q))}{dq} \\ = \; &\frac{\mu}{m} q^{n-1-m} (1-q)^{m-1} \left[n'(n'-m) \binom{n'-1}{m-1} q^{n'-n} - n(n-m) \binom{n-1}{m-1} \right]. \end{aligned} \tag{3.64}$$

When $0 < q < 1$, the sign of (3.64) is determined by the sign of the whole term in the square bracket of (3.64). This whole term is negative when $q = 0$ and positive when $q = 1$, and it is strictly increasing in q. Hence, there exists a unique q^* such that $F^{-1}_{(n',m)}(q) - F^{-1}_{(n,m)}(q)$ is strictly decreasing in q when $0 < q < q^*$ and strictly increasing in q when $q^* < q < 1$. As $F^{-1}_{(n',m)}(0) = F^{-1}_{(n,m)}(0) = 0$, it turns out that when $0 < q \le q^*$, $F^{-1}_{(n',m)}(q) - F^{-1}_{(n,m)}(q) < 0$. As the upper bound of the support of $F_{(n',m)}$ is $\frac{n'\mu}{m}$ while that of $F_{(n,m)}$ is $\frac{n\mu}{m}$, $F^{-1}_{(n',m)}(1) - F^{-1}_{(n,m)}(1) = \frac{n'\mu}{m} - \frac{n\mu}{m} > 0$. As $F^{-1}_{(n',m)}(q) - F^{-1}_{(n,m)}(q)$ is strictly increasing in q when $q^* < q < 1$, there thus exists a unique $\hat{q} \in (q^*, 1)$ such that $F^{-1}_{(n',m)}(\hat{q}) - F^{-1}_{(n,m)}(\hat{q}) = 0$ and $F^{-1}_{(n',m)}(q) < F^{-1}_{(n,m)}(q)$ if and only if $q < \hat{q}$.

Next we prove (b). By Proposition 3.3,

$$F^{-1}_{(n,m')}(q) - F^{-1}_{(n,m)}(q) = \frac{n\mu}{m'} \sum_{i=n-m'}^{n-1} \binom{n-1}{i} q^i (1-q)^{(n-1)-i} - \frac{n\mu}{m} \sum_{i=n-m}^{n-1} \binom{n-1}{i} q^i (1-q)^{(n-1)-i}. \tag{3.65}$$

Differentiating equation (3.65) with respect to q gives

$$\begin{aligned} &\frac{d(F^{-1}_{(n,m')}(q) - F^{-1}_{(n,m)}(q))}{dq} \\ &= n\mu q^{n-1-m}(1-q)^{m'-1} \quad \left[\tfrac{n-m'}{m'}\tbinom{n-1}{n-m'} q^{m-m'} - \tfrac{n-m}{m}\tbinom{n-1}{n-m}(1-q)^{m-m'}\right]. \end{aligned} \tag{3.66}$$

When $0 < q < 1$, the sign of (3.66) is determined by the sign of the whole term in the square bracket of (3.66). This whole term is negative when $q = 0$ and positive when $q = 1$, and it is strictly increasing in q. Hence, there exists a unique q^o such that $F^{-1}_{(n,m')}(q) - F^{-1}_{(n,m)}(q)$ is strictly decreasing when $0 < q < q^o$ and strictly increasing when $q^o < q < 1$. As $F^{-1}_{(n,m')}(0) = F^{-1}_{(n,m)}(0) = 0$, it turns out that when $0 < q \leq q^o$, $F^{-1}_{(n,m')}(q) - F^{-1}_{(n,m)}(q) < 0$. As the upper bound of the support of $F_{(n,m')}$ is $\frac{n\mu}{m'}$ while that of $F_{(n,m)}$ is $\frac{n\mu}{m}$, $F^{-1}_{(n,m')}(1) - F^{-1}_{(n,m)}(1) = \frac{n\mu}{m'} - \frac{n\mu}{m} > 0$. As $F^{-1}_{(n,m')}(q) - F^{-1}_{(n,m)}(q)$ is strictly increasing in q when $q^o < q < 1$, there thus exists a unique $\bar{q} \in (q^o, 1)$ such that $F^{-1}_{(n,m')}(\bar{q}) - F^{-1}_{(n,m)}(\bar{q}) = 0$ and $F^{-1}_{(n,m')}(q) < F^{-1}_{(n,m)}(q)$ if and only if $q < \bar{q}$. $\square$

Proof of Proposition 3.8:

$$\Phi(w, n, m) := \sum_{i=n-m}^{n-1} \binom{n-1}{i} w^i (1-w)^{n-1-i}. \tag{3.67}$$

Note that for any sequence of natural numbers (n_j, m_j) where $n_j \to \infty$ and for all j, $m_j = \rho n_j$, where $0 < \rho < 1$, the weak convergence of quantile function implies that the sequence of distribution functions

$$j \to \Phi(\cdot, n_j, m_j) \tag{3.68}$$

converges weakly to the distribution function of a degenerate random variable concentrated at $1 - \rho$. Thus

$$\lim_{j->\infty} \Phi(w, n_j, m_j) = \begin{cases} 0 & \text{if } w < 1 - \rho \\ 1 & \text{if } w > 1 - \rho \end{cases} \tag{3.69}$$

Let F_Z^j be a sequence of equilibrium distributions for the normalized performance levels then we can express the necessary condition for an equilibrium given by Proposition 3.3 as

$$z = \Phi(F_Z^j(z), n_j, m_j). \tag{3.70}$$

Suppose the proposition is incorrect. Then some sequence $(F_Z^j)_j$ does not converge to the distribution

$$F_Z^\infty(z) = \begin{cases} 1 - \rho & z < 1 \\ 1 & z \geq 1 \end{cases} \tag{3.71}$$

This implies, passing to a subsequence if necessary using Helly's selection theorem (See Billingsley 1985, Theorem 25.9), there exists a limit distribution for the sequence, F'_Z unequal to F_Z^∞. This implies that at some continuity point of F'_Z, $z' \in (0, 1)$,

$$F_Z^j(z') \to F'_Z[z'] \neq 1 - \rho.$$

Exploiting the continuity of Φ, (3.70), and (3.69) and taking limits yield

$$z' = \begin{cases} 0 & \text{if } F'_Z[z'] < 1-\rho \\ 1 & \text{if } F'_Z[z'] > 1-\rho \end{cases} \tag{3.72}$$

Since by hypothesis, $z' \in (0,1)$ and $F[z'] \neq 1-\rho$, equation (3.72) is a contradiction and thus the normalized distribution converges weakly to F_Z^{∞}. Rescaling then implies that the equilibrium distribution converges to F_{∞}. $\square$

Proof of Proposition 3.9: Let $\alpha_S, \alpha_W, \beta_S, \beta_W$ be the nonnegative constants determined by solving the dual problem for types S and W. Note that ψ, defined by equation (3.31), is an increasing concave function defined over the nonnegative real line. Because the two support lines both bound P,

$$\forall x \geq 0,\ \ P(x) \leq \psi(x). \tag{3.73}$$

Because the support of P is contained by the support of the best replies of types W and S, the support of P is contained within

$$\{x \geq 0 : P(x) = \psi(x)\}. \tag{3.74}$$

Thus the conditions of Lemma 3.1 are satisfied and the result follows. $\square$

Proof of Lemma 3.5: First note that it must be the case that $\alpha_W \leq \alpha_S$. This follows from the following argument by contradiction. Suppose that $\alpha_W > \alpha_S$. Then because one of the two support lines cannot lie strictly above the other, it would have to be the case that $\beta_S > \beta_W$. By (3.30) and Proposition 3.9, the support of S's distribution, Supp_S and the support of W's distribution, Supp_W, would satisfy:

$$\begin{aligned} \mathrm{Supp}_S &\subset \{x \geq 0 : \alpha_S + \beta_S\, x \leq \alpha_W + \beta_W\, x\} \\ \mathrm{Supp}_W &\subset \{x \geq 0 : \alpha_S + \beta_S\, x \geq \alpha_W + \beta_W\, x\}. \end{aligned} \tag{3.75}$$

If $\alpha_W > \alpha_S$ and $\beta_S > \beta_W$, equation (3.75) would imply that all points in Supp_S are weakly smaller than all points in Supp_W, which contradicts the mean of S's distribution exceeding the mean of W's distribution. Thus, $\alpha_W \leq \alpha_S$. In order for the S-support line to be a support line, it must not lie strictly above the W-support line, thus $\beta_S \leq \beta_W$. Finally, because by Proposition 3.9, $P(0) = \psi(0)$ and by Lemma 3.4, $P(0) = 0$, it must be the case that $\psi(0) = 0$. By the definition of ψ given by expression (3.31), $\psi(0) = 0$ implies that $\alpha_W = 0$. $\beta_S > 0$ follows because the mean constraint on distributions is always binding. $\square$

Proof of Lemma 3.7: In concession configuration, the weak and the strong's best replies to P lie on disjoint supports and for this configuration to sustain in equilibrium, P must be weakly concave, which is guaranteed if and only if a weak contestant weakly prefers conceding victory to the strong rather than playing the alternative strategy of placing all the mass on 0 with probability $(1 - \frac{\mu_W}{\mu_S})$ and mimicking the strong's distribution with probability $\frac{\mu_W}{\mu_S}$. By playing the prescribed alternative strategy, a weak contestant's probability of winning is $\frac{\mu_W p_S^C}{\mu_S}$, while by conceding victory to the strong, a weak contestant's probability of winning is p_W^C. Thus concession configuration sustains in equilibrium if and only if

$$p_W^C \geq \frac{\mu_W p_S^C}{\mu_S}. \tag{3.76}$$

As $\mu_W/\mu_S = p_W^G/p_S^G$, inequality (3.76) implies that

$$p_W^C/p_S^C \geq p_W^G/p_S^G. \tag{3.77}$$

As ex ante each contestant has the same probability of winning, $\theta p_S^C + (1-\theta)p_W^C = \theta p_S^G + (1-\theta)p_W^G = m/n$. Thus (3.77) is guaranteed if and only if $p_W^C \geq p_W^G$.

In any symmetric Nash equilibrium, weak contestants choose the same distribution F_W and thus have the same chance of winning when matched against each other. Note that this guarantees that by playing the same distribution as other weak contestants, a given weak contestant has at least p_W^C chance of winning. In challenge configuration, the upper bound of the support of a weak contestant's distribution lies strictly above the lower bound of the support of the strong's, suggesting that apart from p_W^C, a weak contestant has some additional chance of winning coming from the strictly positive probability of outperforming strong contestants. Thus for challenge configuration to sustain in a symmetric equilibrium, a weak contestant's probability of winning under the challenge configuration, p_W^G has to be strictly greater than p_W^C. $\square$

Proof of Proposition 3.10: Π^* can be calculated as follows: With probability $\binom{n}{i}\theta^i(1-\theta)^{n-1-i}$ there are i strong contestants in the pool, and in such a case, selection efficiency reaches its maximum when the number of selected strong contestants equals $\min\{i, m\}$. Adding up the expected numbers from $i = 0$ to $i = n$ and dividing the sum by m gives the maximum expected number of selected strong contestants *per winner*, which is just Π^*. The rest of the proof is clear from the definition of $\Delta\Pi$ and the discussion before Proposition 3.10. $\square$

Proof of Corollary 3.3: We first prove (i). It is obvious that when strong contestants have priority over weak ones to win, selection efficiency must be higher than that from a random draw. Thus, $\Pi^* > \theta$. Note that Π^* is independent of r, so $\lim_{r\to 1} \frac{\theta r}{(1-\theta)+\theta r} = \theta < \Pi^*$. Moreover, $\lim_{r\to\infty} \frac{\theta r}{(1-\theta)+\theta r} = 1 > \Pi^*$. As $\frac{\theta r}{(1-\theta)+\theta r}$ is strictly increasing in r for all $r > 1$, there must exist a unique r^* such that $\frac{\theta r^*}{(1-\theta)+\theta r^*} = \Pi^*$. When $1 < r < r^*$, $\Pi = \frac{\theta r}{(1-\theta)+\theta r}$, which is strictly increasing in r and $\Delta\Pi = \Pi^* - \frac{\theta r}{(1-\theta)+\theta r}$, which is positive and strictly decreasing in r. When $r \geq r^*$, $\Pi = \Pi^*$, which is constant in r, and $\Delta\Pi = 0$.

Next we prove (ii). Note that $\partial\Pi^*/\partial\theta$ equals $\sum_{i=0}^{m-1} \frac{n-i}{m}\binom{n}{i}(1-\theta)^{n-i-1}\theta^i$, which is positive and goes to zero when θ goes to one, while $\partial\left[\frac{\theta r}{(1-\theta)+\theta r}\right]/\partial\theta$ equals $\frac{r}{[(1-\theta)+\theta r]^2}$, which is positive and goes to $1/r$ when θ goes to one. Thus, both Π^* and $\frac{\theta r}{(1-\theta)+\theta r}$ are strictly increasing in θ and the rate of increase is smaller for Π^* than for $\frac{\theta r}{(1-\theta)+\theta r}$ when θ approaches one. These imply two facts. First, Π, which equals $\min\left\{\frac{\theta r}{(1-\theta)+\theta r}, \Pi^*\right\}$, must be strictly increasing in θ. Second, as when $\theta = 1$, $\Pi^* = \frac{\theta r}{(1-\theta)+\theta r} = 1$, for r sufficiently close to 1, $\Pi^* > \frac{\theta r}{(1-\theta)+\theta r}$ and thus $\Delta\Pi > 0$. As $\Delta\Pi = 0$ for $\theta = 0$ and for $\theta = 1$ and as $\Delta\Pi$ is continuous in θ, it must be nonmonotonic in θ and there must exist $\hat{\theta} \in (0, 1)$ at which $\Delta\Pi$ is maximized.

Now we prove (iii). For each realization of the number of strong contestants in the original pool, adding one more contestant into the pool will never decrease the number of strong contestants but will strictly increase this number in expectation. As Π^* requires selecting no weak contestants before all strong contestants are selected, it is obvious that Π^* must be strictly increasing with one more contestant added into the pool and thus strictly increasing in n. Moreover, when n goes to infinity, by the law of large numbers, the proportion of strong contestants goes to θ, implying that the number of strong contestants goes to infinity, so Π^* must go to 1, which is strictly greater than $\frac{\theta r}{(1-\theta)+\theta r}$. When $n = m$, $\Pi^* = \theta$, which is strictly less than $\frac{\theta r}{(1-\theta)+\theta r}$. As $\frac{\theta r}{(1-\theta)+\theta r}$ is constant in n, there must exist a threshold n^* such that when $n \leq n^*$, $\Pi = \Pi^*$, which is strictly increasing in n, and when $n > n^*$, $\Pi = \frac{\theta r}{(1-\theta)+\theta r}$, which is constant in n. The result on $\Delta\Pi$ follows directly from (3.47).

To prove (iv), by equation (3.45), it is obvious that Π^* is strictly decreasing in m when m takes integers. Note that when $m = 1$, $\Pi^* = 1 - (1-\theta)^n$. If this value is weakly less than $\frac{\theta r}{(1-\theta)+\theta r}$, which condition is equivalent to (3.48), $\Pi^* \leq \frac{\theta r}{(1-\theta)+\theta r}$ for all $m \geq 1$, and thus in this case $\Pi = \Pi^*$ and $\Delta\Pi = 0$ for all $m \geq 1$. If (3.48) is violated, then when $m = 1$, $\Pi^* > \frac{\theta r}{(1-\theta)+\theta r}$, while when $m = n$, $\Pi^* = \theta < \frac{\theta r}{(1-\theta)+\theta r}$, so there must exist a threshold m^* such that when $m < m^*$, $\Pi = \frac{\theta r}{(1-\theta)+\theta r}$, which is constant in m, and when $m^* \leq m \leq n$, $\Pi = \Pi^*$, which is strictly decreasing in m. The result on $\Delta\Pi$ follows directly from (3.47).

To prove (v), note that by Proposition 3.10, changes in n and m can affect Π and $\Delta\Pi$ only by affecting Π^*. As both Π and $\Delta\Pi$ are weakly increasing in Π^*, to prove (v), all we need to show is the following: Whenever k is an integer and $k > 1$,

$$\Pi^*(k\,n, k\,m) \geq \Pi^*(n, m). \tag{3.78}$$

To see this, let $\tilde{S}_q$ represent the sum of q iid Bernoulli random variables. Let $\tilde{A}_q = \tilde{S}_q/q$ represent the average of q iid Bernoulli random variables. Consider Π^* as a function of n and m. Then we can express Π^* as

$$\Pi^*(n, m) = \mathbb{E}\left[\min\left[\frac{\tilde{S}_n}{m}, 1\right]\right]. \tag{3.79}$$

Similarly the maximum selection efficiency after a scale increase will be given by

$$\Pi^*(k\,n, k\,m) = \mathbb{E}\left[\min\left[\frac{\tilde{S}_{k\,n}}{k\,m}, 1\right]\right]. \tag{3.80}$$

Note that we can write

$$\frac{\tilde{S}_{k\,n}}{k\,m} = \frac{k-1}{k}\left(\frac{n}{m}\,\frac{\tilde{S}_{(k-1)\,n}}{(k-1)\,n}\right) + \frac{1}{k}\left(\frac{n}{m}\,\frac{\tilde{S}_n}{n}\right) = \frac{k-1}{k}\left(\frac{n}{m}\,\tilde{A}_{n(k-1)}\right) + \frac{1}{k}\left(\frac{n}{m}\,\tilde{A}_n\right). \tag{3.81}$$

Because the function $x \to \min[x, 1]$ is concave,

$$\mathbb{E}\left[\min\left[\frac{\tilde{S}_{k\,n}}{k\,m}, 1\right]\right] \geq \frac{k-1}{k}\,\mathbb{E}\left[\min\left[\left(\frac{n}{m}\,\tilde{A}_{n(k-1)}\right), 1\right]\right] + \frac{1}{k}\,\mathbb{E}\left[\min\left[\left(\frac{n}{m}\,\tilde{A}_n\right), 1\right]\right]. \tag{3.82}$$

Because $n(k-1) \geq n$ and because averages of iid random variables which include more terms in the average weakly second-order stochastically dominate averages with fewer terms, and because $x \to \min[x, 1]$ is concave and non-decreasing,

$$\mathbb{E}\left[\min\left[\left(\frac{n}{m}\,\tilde{A}_{n(k-1)}\right), 1\right]\right] \geq \mathbb{E}\left[\min\left[\left(\frac{n}{m}\,\tilde{A}_n\right), 1\right]\right]. \tag{3.83}$$

Therefore, inequalities (3.82) and (3.83) imply that

$$\mathbb{E}\left[\min\left[\frac{\tilde{S}_{k\,n}}{k\,m}, 1\right]\right] \geq \mathbb{E}\left[\min\left[\left(\frac{n}{m}\,\tilde{A}_n\right), 1\right]\right] = \mathbb{E}\left[\min\left[\frac{\tilde{S}_n}{m}, 1\right]\right]. \tag{3.84}$$

Thus, (3.84), (3.80), and (3.79) establish (3.78). $\square$

Proof of Proposition 3.11: With the imposition of a scoring cap $\bar{x}$, each contestant's problem becomes choosing probability measures over $[0, \bar{x}]$ instead of over $[0, \infty)$ to maximize his chance of winning. On $[0, \bar{x}]$, equations (3.29) and (3.30) still hold (the discussion can be referred to Section 3.2). Thus the relationships between Lagrange multipliers still satisfy Lemma 3.5, which suggests that there are still two equilibrium configuration candidates: concession and challenge. Which configuration is the equilibrium configuration is still presented by Lemma 3.7 whose proof is not affected by the imposition of $\bar{x}$. It is obvious that p_W^C and p_S^C are unaffected by $\bar{x}$ as they are determined by the prize allocation rule that strong contestants have priority of winning over weak contestants which is independent of $\bar{x}$. Note that as long as $\bar{x} \geq \mu_S$, each type's probability of winning conditional on challenge configuration, which is given by equation (3.44), is also unaffected by $\bar{x}$. Thus, by Lemma 3.7, equilibrium configuration is unaffected and therefore each type's equilibrium probability of winning is unaffected. By Bayes' Rule, selection efficiency is thus unaffected. $\square$

Proof of Proposition 3.12: We start our analysis by first arguing that imposing $\bar{x}$ does not affect W-support line provided that $\bar{x} \geq \mu_S$, nor does it affect S-support line provided that $\bar{x} > \mu_S$. Note that when $\bar{x} = \mu_S$, S-support line is indeterminate, but this brings no difficulty to our analysis as we know that strong contestants place all mass at the upper bound when $\bar{x} = \mu_S$. We denote by $\alpha_{t,\bar{x}}$ and $\beta_{t,\bar{x}}$ the multipliers from this modified dual problem for type-t contestants, for all $t \in \{S, W\}$.

Implied by Lemma 3.8, $P(0) = 0$, so the concave lower envelope of both types' upper support lines intersects the origin, implying that $\alpha_{W,\bar{x}} = 0$. Note that $\alpha_{W,\bar{x}} + \beta_{W,\bar{x}}\mu_W = p_W$.

As both $\alpha_{W,\bar{x}}$ and p_W are constant in $\bar{x}$ provided that $\bar{x} \geq \mu_S$, $\beta_{W,\bar{x}}$ must be constant in $\bar{x}$ as long as $\bar{x} \geq \mu_S$. This proves that W-support line is unaffected by the imposition of $\bar{x}$ when $\bar{x} \geq \mu_S$. Because imposing $\bar{x} \geq \mu_S$ does not change equilibrium configuration, if challenge equilibria exist in the contest without $\bar{x}$, S-support line, as it overlaps W-support line in challenge configuration, must remain unchanged with the imposition of $\bar{x} > \mu_S$. Thus, below we only have to show that S-support line is also unaffected by the imposition of $\bar{x} > \mu_S$ if the contest without $\bar{x}$ has concession equilibria.

Note that $\alpha_{S,\bar{x}}$ and $\beta_{S,\bar{x}}$ in concession equilibria are determined by the following two equations: (1) $\alpha_{S,\bar{x}} + \beta_{S,\bar{x}}\mu_S = p_S^C$ and (2) $\beta_{W,\bar{x}}\check{x} = \check{p}$, where $\check{x}$ represents the performance level at which S-support line and W-support line intersect, i.e., $\beta_{W,\bar{x}}\check{x} = \alpha_{S,\bar{x}}+\beta_{S,\bar{x}}\check{x}$, and $\check{p}$ equals a contestant's probability of winning if he concedes victory to all strong contestants but has priority over all weak ones to win. The second equation must hold as P must meet the concave lower envelope of the two upper support lines at $\check{x}$, since otherwise P has to break contact with the concave lower envelope at a point smaller than $\check{x}$ and can only jump back onto the concave lower envelope at $\bar{x}$, implying that strong contestants put all mass at $\bar{x} > \mu_S$, which violates a strong contestant's capacity constraint. As $\bar{x}$ plays no role in both equations, $\alpha_{S,\bar{x}}$ and $\beta_{S,\bar{x}}$ are constant in $\bar{x}$ so long as $\bar{x} > \mu_S$. Thus, S-support line is unaffected by $\bar{x} > \mu_S$ if the contest without $\bar{x}$ has concession equilibria.

The above analysis shows that the concave lower envelope of the two upper support lines is unaffected by the imposition of $\bar{x}$ provided that $\bar{x} > \mu_S$. Thus, when $\bar{x} \geq \hat{x}$, where $\hat{x}$ is defined in Proposition 3.9, equilibrium distributions are unaffected. When $\mu_S \leq \bar{x} < \hat{x}$, $P(\bar{x}) < P(\hat{x}) = 1$. This implies that there must be point mass at $\bar{x}$, which further implies discontinuity of P at $\bar{x}$. In this case, P must break contact with the concave lower envelope at some point $\tilde{x}$, with $0 < \tilde{x} < \bar{x}$. Then by Lemma 3.1, P traces out the concave lower envelope on $[0, \tilde{x}]$, keeps flat on $[\tilde{x}, \bar{x})$, and jumps back onto the concave lower envelope at $\bar{x}$. Thus, the imposition of $\mu_S \leq \bar{x} < \hat{x}$ induces each contestant to transfer the mass over $[\tilde{x}, \hat{x}]$ to the point mass at $\bar{x}$, leaving the mean of his distribution and the mass below $\tilde{x}$ unchanged. Such a change in distribution satisfies the conditions for a mean-preserving contraction defined by Diamond and Stiglitz (1974) and Machina and Pratt (1997), who show that mean-preserving contraction is preferred by every risk-averse utility maximizer. $\square$

Proof of Proposition 3.14: Let k denote the number of local contests and let n_i and m_i denote the number of contestants and the number of prizes in the ith local contest respectively for all $1 \leq i \leq k$. n_i and m_i must satisfy $\sum_{i=1}^{k} n_i = n$ and $\sum_{i=1}^{k} m_i = m$. Denote by $\Pi(n_i, m_i)$ and $\Pi^*(n_i, m_i)$ actual selection efficiency and maximum selection efficiency of the ith local contest respectively and by Π^g and Π^l actual selection efficiency of the grand contest structure and of the local contest structure respectively. As assigning contestants to a local contest with zero prize is equivalent to reducing the number of contestants, which never improves actual selection efficiency, below we focus on the case where $m_i \geq 1$ for all $1 \leq i \leq k$.

First, we prove that localizing a contest never improves selection efficiency. This is equivalent to show that $\Pi^g \geq \Pi^l$. Note that the grand contest is the efficient selection mechanism if it has concession equilibria, in which case $\Pi^g \geq \Pi^k$. If the grand contest has challenge equilibria, implied by Proposition 3.10, $\Pi^g = \frac{\theta r}{(1-\theta)+\theta r}$. By Proposition 3.10,

$$\Pi^l = \sum_{i=1}^{k} \left(\frac{m_i}{m}\right) \min\left\{\frac{\theta r}{(1-\theta)+\theta r}, \Pi^*(n_i, m_i)\right\}, \tag{3.85}$$

which is bounded by $\frac{\theta r}{(1-\theta)+\theta r}$ from above. Thus $\Pi^g \geq \Pi^l$ if the grand contest has challenge equilibria. Therefore, $\Pi^g \geq \Pi^l$ regardless of the equilibrium configuration of the grand contest.

Next, we prove that challenge equilibria in every local contest is sufficient for $\Pi^l = \Pi^g$. Note that if every local contest has challenge equilibria, $\Pi^l = \frac{\theta r}{(1-\theta)+\theta r}$. Implied by Proposition 3.10, Π^g is bounded by $\frac{\theta r}{(1-\theta)+\theta r}$ from above, so $\Pi^g \leq \Pi^l$ in this case. As we have proved above that $\Pi^g \geq \Pi^l$, it must be that $\Pi^g = \Pi^l$.

Finally, we prove that challenge equilibria in every local contest is necessary for $\Pi^l = \Pi^g$. If there exists at least one local contest with concession equilibria, $\Pi^l < \frac{\theta r}{(1-\theta)+\theta r}$. In this case, if the grand contest has challenge equilibria, $\Pi^g = \frac{\theta r}{(1-\theta)+\theta r} > \Pi^l$; if the grand contest has concession equilibria, $\Pi^g = \Pi^*(n,m)$, in which case we can still show that $\Pi^g > \Pi^l$ by equation (3.85) and the following result:

$$\Pi^*(n,m) > \sum_{i=1}^{k} \left(\frac{m_i}{m}\right) \Pi^*(n_i, m_i). \tag{3.86}$$

To prove equation (3.86), all we need to show is that a local contest mechanism is not an efficient selection mechanism even if all local contests have concession equilibria. The reason is as follows. As $m < n$, under a local contest structure, there must exist a local contest, assuming it to be the jth local contest without loss of generality, in which $m_j < n_j$. As every local contest has at least one prize, $m_j < m$. With a positive probability, there are $(m_j + 1)$ strong contestants among n contestants. As $m_j + 1 \leq m$, an efficient selection requires selecting all $(m_j + 1)$ strong contestants. However, under a local contest structure, with a positive probability, the $(m_j + 1)$ strong contestants are all assigned to the jth local contest, in which case it is impossible for all of them to be selected. This proves (3.86) and completes the proof of necessity. $\square$

Bibliography

[1] Akerlof, Robert J. and Richard T. Holden, 2012, The nature of tournaments, *Economic Theory* 51, 289-313.

[2] Billingsley, Patrick, 1985, *Probability and measure* (John Wiley and Sons).

[3] Brown, Keith C., W. V. Harlow, and Laura T. Starks, 1996, Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, *The Journal of Finance* 51, 85-110.

[4] Chevalier, Judith and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167-200.

[5] Clark, Derek J. and Christian Riis, 2001, Rank-order tournaments and selection, *Journal of Economic* 2, 167-91.

[6] Diamond, Peter A. and Joseph E. Stiglitz, Joseph E., 1974, Increases in risk and in risk aversion, *Journal of Economic Theory* 8, 337-60.

[7] Dijk, Oege, Martin Holmèn, and Michael Kirchler, 2011, Rank matters: The impact of social competition on portfolio choice, Working paper.

[8] Dubins, L. E. and L. J. Savage, 1965, *How to gamble if you must. Inequalities for stochastic processes* (McGraw-Hill).

[9] Falkenstein, Eric G., 1996, Preferences for stock characteristics as revealed by mutual fund portfolio holdings, *Journal of Finance* 51, 111-35.

[10] Fu, Qiang and Jingfeng Lu, 2009, The beauty of "bigness": On optimal design of multi-winner contests, *Games and Economic Behavior* 66, 146-61.

[11] Gaba, Anil, Ilia Tsetlin, Robert L. Winkler, 2004, Modifying variability and correlations in winner-take-all contests, *Operations Research* 52, 384-95.

[12] Gilpatric, Scott M., 2009, Risk taking in contests and the role of carrots and sticks, *Economic Inquiry* 47, 266-77.

[13] Grote, Dick, 2002, Forced ranking: Behind the scenes, *The Conference Board Revew.*

[14] Hart, Sergiu, 2008, Discrete Colonel Blotto and General Lotto games, *International Journal of Game Theory* 36, 441-60.

[15] Hosking, J. R. M., 1990, L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society. Series B (Methodological)* 52, 105-24.

[16] Hvide, Hans K., 2002, Tournament rewards and risk taking, *Journal of Labor Economics* 20, 877-98.

[17] Hvide, Hans K. and Eirik G. Kristiansen, 2003, Risk taking in selection contests, *Games and Economic Behavior* 42, 172-79.

[18] Jones, Chris M., 2002, The complementary beta distribution, *Journal of Statistical Planning and Inference* 104, 329-37.

[19] Kempf, Alexander, Stefan Ruenzi, and Tanja Thiele, 2009, Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry, *Journal of Financial Economics* 92, 92-108.

[20] Kräkel, Matthias, 2008, Optimal risk taking in an uneven tournament game with risk averse players, *Journal of Mathematical Economics* 44, 1219-31.

[21] Lazear, Edward P. and Sherwin Rosen, 1981, Rank-order tournaments as optimum labor contracts, *Journal of Political Economy* 89, 841-64.

[22] Lin, Jerchern, 2011, Fund convexity and tail risk-taking, Working paper.

[23] Lizzeri, Alessandro, 1999, Budget deficits and redistributive politics, *Review of Economic Studies* 66, 909-28.

[24] Lizzeri, Alessandro and Nicola Persico, 2001, The provision of public goods under alternative electoral incentives, *American Economic Review* 91, 225-39.

[25] Machina, Mark J. and John W. Pratt, 1997, Increasing risk: Some direct constructions, *Journal of Risk and Uncertainty* 14, 103-27.

[26] Meyer, Margaret A., 1991, Learning from coarse information: Biased contests and career profiles, *Review of Economic Studies* 58, 15-41.

[27] Moldovanu, Benny and Aner Sela, 2006, Contest architecture, *Journal of Economic Theory* 126, 70-96.

[28] Moldovanu, Benny, Aner Sela, Xianwen Shi, 2010, Carrots and sticks: Prizes and punishments in contests, *Economic Inquiry* 50, 453-62.

[29] Myerson, Roger B., 1993, Incentives to cultivate favored minorities under alternative electoral systems, *American Political Science Review* 87, 856-69.

[30] Nanda, Asok and Moshe Shaked, 2001, The hazard rate and the reversed hazard rate orders, with applications to order statistics, *Annals of the Institute of Statistical Mathematics* 53, 853-64.

[31] Nieken, Petra, 2010, On the choice of risk and effort in tournaments - Experimental evidence, *Journal of Economics & Management Strategy* 19, 811-40.

[32] Nieken, Petra and Dirk Sliwka, 2010, Risk-taking tournaments - Theory and experimental evidence, *Journal of Economic Psychology* 31, 254-68.

[33] Oja, Hannu, 1981, On location, scale, skewness and kurtosis of univariate distributions, *Scandinavian Journal of Statistics* 8, 154-68.

[34] Pyle, David H. and Stephen J. Turnovsky, 1970, Safety-first and expected utility maximization in mean-standard deviation portfolio analysis, *The Review of Economics and Statistics* 52, 75-81.

[35] Reiss, R.-D., 1980, *Approximate distributions of order statistics* (Springer Verlag).

[36] Roberson, Brian, 2006, The Colonel Blotto game, *Economic Theory* 29, 1-24.

[37] Rose-Ackerman, Susan, 1991, Risk taking and ruin: Bankruptcy and investment choice, *The Journal of Legal Studies* 20, 277-310.

[38] Ryvkin, Dmitry, 2010, The selection efficiency of tournaments, *European Journal of Operational Research* 206, 667-75.

[39] Ryvkin, Dmitry and Andreas Ortmann, 2008, The predictive power of three prominent tournament formats, *Management Science* 54, 492-504.

[40] Sahuguet, Nicolas and Nicola Persico, 2006, Campaign spending regulation in a model of redistributive politics, *Economic Theory* 28, 95-124.

[41] Sato, Michikazu, 1997, Some remarks on the mean, median, mode and skewness, *Australian Journal of Statistics* 39, 219–24.

[42] Taylor, Jonathan, 2003, Risk-taking behavior in mutual fund tournaments, *Journal of Economic Behavior & Organization* 50, 373-83.

[43] Wagner, Niklas and Elisabeth Winter, 2013, A new family of equity style indices and mutual fund performance: Do liquidity and idiosyncratic risk matter? *Journal of Empirical Finance* 21, 69-85.