Economies of Extremes:
Lessons from Venture-Capital Decision Making

Suzanne de Treville[a,*], Jeffrey Petty[a], Stefan Wager[b]

[a]University of Lausanne, Faculty of Business and Economics
1015 Lausanne, Switzerland
[b]Stanford University, Department of Statistics
Stanford, CA – 94305, U.S.A.

Abstract

An organization's ability to exploit extreme events—such as exceptional opportunities—depends on its capacity strategy. The venture capital industry illustrates the interplay of expensive capacity and negative externalities from high utilization. The cost of adding a venture capitalist provides a strong incentive to run lean, but such leanness may make it impossible to evaluate all interesting investment opportunities. Using concepts from extreme-value theory, we analyze the trade-off between the costs and benefits arising from an increase in the number of evaluated deals. We ground our analysis in 11 years of archival data from a venture capital firm, representing 3,631 deals, the decisions made, the reasons for those decisions, and the decision lead times. The firm identified 20% of arriving deals as worth evaluating during the screening process, but was not able to evaluate approximately 9% of those interesting deals due to a lack of capacity. We show that the value of increasing the number of deals evaluated increases with the tail weight of the distribution of deal values. When the right tail is light, increasing the number of deals evaluated may provide too modest a benefit to justify the cost. When, however, the right tail is heavy, the value of increasing the number of deals is likely to more than compensate for the cost of capacity. Our results provide new insight into the relative value of a chase capacity strategy that emphasizes responsiveness versus a high-utilization heuristic that emphasizes productivity. Our approach can be applied to other search operations such as personnel selection, quality circles seeking to identify root causes, and making employee capacity available for innovation.

*Corresponding author: suzanne.detreville@unil.ch; Tel.:+41 21 692 3448

1. Introduction

In a survey of literature on strategic capacity management, Van Mieghem (2003) described the tension arising from the interplay of expensive capacity and negative externalities from high utilization, noting that unmet demand is likely when capacity is expensive. In this paper, we use a search operation to study how extreme values shape this interplay. Our analysis is grounded in 11 years of historical data concerning capacity shortages occurring in a venture-capital firm. During the two funds covered by the data, the firm followed a policy of only adding capacity when absolutely unavoidable because of the high cost of hiring an additional venture capitalist. When a deal arrived to the venture-capital firm, an initial screening selected those deals deemed interesting enough for a thorough evaluation. Of deals recorded by the decision maker as interesting enough for evaluation, 9% were nevertheless rejected, with the reason cited being lack of capacity. The historical data thus suggests a type of level capacity strategy in which demand was foregone when the maximum service rate was exceeded. The research question that arose from the historical data was whether the savings from postponing an additional hire—running lean—compensated for the cost of not evaluating all interesting deals.

The increase in fund value arising from an increase in capacity allocated to evaluation depends on many factors, including the quality of the additional deals evaluated, the competence of the venture capitalists in choosing deals, the dynamics of the deal-arrival process, and the heaviness of the right tail of the distribution of deal values. Our focus is the right-tail weight. A venture-capital firm invests in a small number of deals relative to the sample from which they are selected: The firm described here received 3,631 deals, evaluated 700, and selected what they considered to be the best 35 of those 700. Had they evaluated more than 700 deals, some deals at the bottom of the top 35 out of 700 probably would have been replaced by better deals. Does swapping out a small number of good deals for better deals increase the overall portfolio value enough to justify the cost of the extra capacity required?

Extreme-value-theory methods emerge as natural tools for addressing this question. As we deepen the search, the value of a portfolio made up of the top 35 deals will increase at the same rate as the sample maxima. We can characterize this rate under regularity conditions. We show that the

relationship between capacity strategy and profit hinges on the magnitude of extremes, as captured by the tail index of the distribution of deal values. The tail index is an extreme-value-theoretic construct that describes the decay rate of the right tail. Distributions with a thin right tail, such as the Gaussian distribution, have a null index. As the tail index approaches 0.5, the variance becomes infinite, and at a tail index of 1 the mean becomes infinite.

To make the problem mathematically tractable, we begin with three simplifying assumptions: 1) the venture capitalists are perfectly competent at evaluating deals, 2) all deals are considered simultaneously, and 3) the additional deals evaluated are drawn from the same distribution as the original deals. We show that for a tail index of γ, an increase in the number of evaluated deals by a factor of α is expected to increase the average value of the portfolio by α^γ. We later demonstrate that this result is robust to violations of the first two simplifying assumptions, and that it is possible to estimate the impact of violating the third assumption.

To gain intuition about the tail weight of deal values that a venture-capital firm might be expected to face, we estimated the tail index of the distribution of returns of initial public offerings (IPOs). If the tail weight of deal values resembles the $\gamma = 0.4$ tail weight of IPO returns, our model predicts that a 9% increase in the number of deals evaluated should produce an expected increase of 3.5% in the value of a portfolio made up of the 35 top deals. Applied to an average venture-capital portfolio of $151 million, a 3.5% increase yields $5 million, more than enough to fund an increase in capacity.

Our results extend beyond the venture-capital context, and contribute to a better understanding of the value of capacity in a search operation. The general insight that increasing sampling by a factor of α is expected to increase the value of the sample maxima by a factor of α^γ can be used to quantify the value of allocating capacity to a search operation. The tail weight has been incorporated as one parameter of the distribution of concept values in balancing cost and benefit of search in product design (Dahan and Mendelson 2001). Tail weight can also play a role in the decision about how much capacity to allocate to search operations such as personnel selection and root-cause identification by quality circles. In Section 6 we describe the application of the model to a decision faced by a large organization about how much capacity to allocate to personnel selection.

More generally, our analysis gives insight into how a positive tail weight impacts the value of

a capacity buffer. Having a capacity buffer—that is, following a chase capacity strategy—allows a firm to take advantage of peak demand. Managers often find it difficult to justify maintaining slack capacity just in case peak demand would occur, especially in light of the lean-production paradigm and its focus on eliminating rather than accommodating variability (e.g., Womack et al. 1990). Firms seeking to maintain a high capacity utilization are more likely to follow a level capacity strategy that allows for stockouts and may entail building inventory during times of low demand. Deploying a level capacity strategy is often done in conjunction with practices aimed to tame demand variability. As extreme-value theory is not commonly used in the operations-management field, managers are typically not equipped to quantify the value of the capacity buffer, and may find it difficult to justify maintaining a capacity buffer even when intuition argues that it would be of high value. Van Mieghem (2003, p. 281) referred to the "continental divide between inventory and queuing models [that] also applies to the subfield of capacity strategy". Incorporation of the tail index into the capacity-strategy literature may contribute to bridging this divide.

2. Relevant Literature

One approach to capacity strategy arises from literature on reducing lead times, with its emphasis on the importance of avoiding excessive resource utilizations (e.g., Hopp and Spearman 2001, Suri 1998). The queuing models common to this literature provide a way of valuing a capacity buffer in the face of demand uncertainty. More generally, slack capacity has been established as a key source of volume flexibility (Jack and Raturi 2002). Hayes and Wheelwright (1984) proposed as a rule of thumb that firms establish a modest capacity buffer (on the order of 10%) in the face of demand variability. Authors such as Kellogg and Nie (1995) and Goodale et al. (2003) have highlighted the need for capacity buffers in service operations facing high arrival-rate variability. Koste et al. (2004), however, observed that volume flexibility was less addressed in the literature than other types of flexibility, possibly indicating some ambivalence in how capacity buffers are viewed in the operations management field.

Van Mieghem (2003, p. 275) explored and summarized what he referred to as the "direct functional dependence of operating profit on the capacity stock" that depends on time, the resources that make

up its "capacity portfolio", and uncertainty. Uncertainty encourages firms to optimally plan demand-capacity imbalances. Van Mieghem (2003) observed, however, that those working to implement such a capacity buffer faced the hurdle of explaining to management the need to invest in capacity that might not be fully utilized.

The emphasis on maintaining high resource utilization is exemplified by the Toyota Production System and the lean-production approach that it spawned. Lean production emphasizes line balancing through practices such as bottleneck elimination and cycle-time reduction (e.g., Shah and Ward 2003). Tools like heijunka combine with line balancing to allow utilization to be maximized (e.g., Womack et al. 1990). The capacity-profit functional relationship under lean production incorporates the idea that high utilization encourages learning, with intentional penury of capacity used to create exploratory stress that encourages productivity improvements (e.g., Monden 2011, Suri and de Treville 1986).

Capacity strategy tends to focus on demand uncertainty. Extending the focus to search operations provides new insight into how to value a capacity buffer. Balancing the cost and value of search is addressed in literature about the innovation process (e.g., Chao et al. 2009, Dahan and Mendelson 2001, Loch and Kavadias 2002, Terwiesch and Loch 2004, Terwiesch and Xu 2008, Weitzman 1979), with selection of deals by venture capitalists given as an example of such a search operation (Terwiesch and Xu 2008, Terwiesch and Ulrich 2009).

Dahan and Mendelson (2001) estimated the optimal number of concepts to test when the distribution of concept values follows a generalized-extreme-value distribution (GEV) with scale parameter b and tail-shape parameter $\alpha = \gamma^{-1}$, and with a per-test cost c. They showed that the optimal number of concepts to test for a null tail index was $\frac{b}{c}$, becoming $\left[\frac{b}{c\alpha}\Gamma(\frac{\alpha-1}{\alpha})\right]^{\frac{\alpha}{\alpha-1}}$ as the tail-shape parameter became positive. For a null tail weight, changes in the per-unit cost of testing were shown to have a constant effect on the optimal amount to spend on testing. As the tail weight became positive, the optimal spending for testing was shown to be convex increasing in decreases in c.

Although many of the researchers applying extreme-value theory to the innovation process have focused primarily on thin-tailed distributions (e.g., Erat and Kavadias 2008, Girotra et al. 2010, Terwiesch and Xu 2008), there is recognition that allowing more flexibility in modeling tails should

be considered when deciding how much capacity to allocate to search. Singh and Fleming (2010, p. 53) stated, "Theorists should increase their metaphorical 'degrees of freedom'; rather than focusing only on the mean or the symmetric variance, or both, of a distribution, they should begin thinking about the different effects a variable may have on the lower tail, the mean, and the upper tail." (See also Fleming 2007). Silverberg and Verspagen (2007) demonstrated empirically that the distributions of innovation success as measured by patent citations and returns were not only skewed, but had a heavy right tail. These authors brought up a seldom-considered but crucial practical issue: A lognormal distribution will often fit a heavily skewed and fat-tailed dataset better than a generalized pareto distribution because of the scarcity of tail data, causing systematic underestimation of the tail heaviness. Silverberg and Verspagen (2007) argued that the strategic importance of incorporating correct tail weight information in decision making justified choosing the distribution to fit the tail rather than the body.
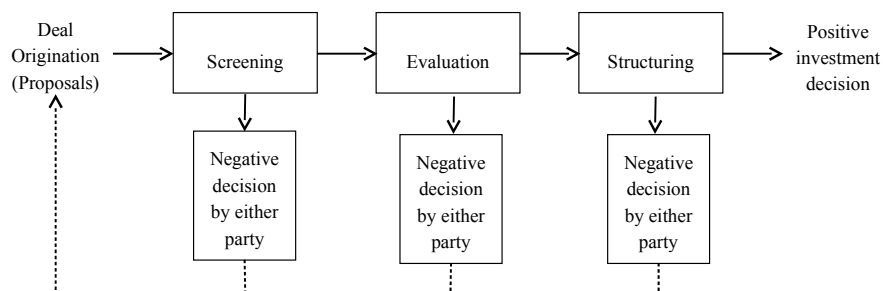
3. The Venture-Capital Firm

3.1. Venture-Capital Operations

Venture-capital firms organize their activities into 10-year cycles. Investors provide capital to the firm at the beginning of the period; the venture-capital firm then gradually evaluates and selects portfolio firms in which to invest. After investment, these portfolio firms are managed by the venture-capital firm until exit. At the end of 10 years, the venture-capital firm exits all remaining deals and returns the proceeds of the fund to the investors. Figure 1 shows the venture-capital investment process in a typical firm (Fried and Hisrich 1994, Hall and Hofer 1993, Tyebjee and Bruno 1984b). Not all decisions are final: A revision of a rejected deal may be resubmitted.

Scarce capacity for all activities carried out by venture capitalists (from screening and evaluation to portfolio management) is a common theme in the venture-capital literature. Shepherd et al. (2005) described the agency problem that emerges as venture capitalists struggle to provide enough attention to portfolio firms and pre-investment activities (see also Kanniainen and Keuschnigg 2003). Resource limitations have been shown to impact venture capitalists' attention and allocation of time across portfolio firms (Gifford 1997, Gorman and Sahlman 1989, Macmillan et al. 1985, Sapienza and Gupta

Figure 1: The Venture Capital Investment Process.

1994). Riquelme and Rickards (1992) observed that scarce capacity encouraged venture capitalists to use intuition rather than analysis in choosing which deals to evaluate.

Venture-capital firms tend to be staffed by a small team of venture capitalists who select the deals and manage the portfolio firms, and also acquire the funds to be invested (Tyebjee and Bruno 1984b). Sahlman (1990) described how relatively few (6 to 12) venture capitalists manage portfolios of sizes up to around $200 million, with an average of $12 to $34 million per professional. Robinson (1987) surveyed the way venture capitalists from 53 venture-capital firms divided their time between deal selection and portfolio management, as well as promotion of the firm, searching for new deals, evaluation of the current portfolio, and administration. Deal selection was observed to consume around a third of venture capitalists' capacity (see also Tyebjee and Bruno 1984b).

3.2. Deal Arrivals

The venture capital firm considered in this study has multiple offices in Europe and invests globally in seed, early, and late-stage companies within a single industry. The firm, of average size for the industry, ranks in the top quartile of European-based venture capital firms in terms of fund performance according to the Thompson VentureXpert data base. The team is comprised of people from four different countries, each holding a graduate degree with at least five years of relevant industry experience. Deal data spans an 11-year period and includes information on 3,631 deals that were submitted to the firm during the life of two venture funds. The same team—consisting on average of 5 venture capitalists—was responsible for the management of both funds. The two funds had some temporal overlap, but origination and screening phases of Fund II were not initiated until after the

Table 1: Action Log

The date the proposal entered the process
Whether the proposal arrived directly or was referred by another party
The stage and location of the company
Ongoing venture-capitalist comments about the deal
The decision taken
Justification for the decision
The date the decision was made

final portfolio of companies in Fund I had been selected. At the time of data collection the firm had made a total of 35 portfolio investments (around 1% of submitted deals) across the two funds, dividing their role as lead or co-investor almost equally across portfolio firms. The acceptance rate is consistent with the 1 to 4% reported in the literature (Maier and Walker 1987, Tyebjee and Bruno 1984a). Deals that were either not accepted by the firm or that the firm was not able to pursue were referred to as "dead" deals.

When a proposal arrived to the firm, a deal file was created that contained company document submissions, email correspondence, and supporting information such as industry reports and opinions from external experts. The firm also maintained an action log (representing 7,284 pages of text) for the information categories listed in Table 1.

One of the authors spent several months at the venture-capital firm to collect and code the data in the action log, also carrying out extensive semi-formal interviews with the venture capitalists and administrative staff. The original focus of the research was the decision process of the venture capitalists, especially the reasons for the decisions that were made. This research is described in detail in *reference removed to protect author identity*. Along the way an interesting and unexpected capacity-strategy issue emerged: The venture-capital firm was running at such high capacity utilization that it routinely chose to discard otherwise interesting deals. When lack of capacity was a major factor in rejecting a deal, this was recorded in the log. Excerpts from these log comments are given in Table 2.

The venture capitalists at the firm estimated that around 20% of deals made it past screening to the evaluation stage, consistent with literature on the venture-capital process (e.g., Roberts 1991); thus, we infer that the firm evaluated around 700 deals.

During the 11-year period covered by our data, the firm recorded 66 instances of capacity-related

Table 2: Reasons for Capacity-Related Rejections

"Rejected due to high activity on other deals at this time"
"Too busy with other projects"
"Currently have no capacity to follow up"
"Team decided not to pursue this deal due to time constraints"
"Not at the moment due to our Deal Flow"
"Business plan reviewed, though interesting, has to take less priority to other projects"
"Interesting but time constraints due to other due diligence"
"Interesting technology but DEAD due to time constraints with other deals"
"Four weeks are unrealistic for us given the early stage of our process"
"Although not uninteresting - lack of time to go further"
"Due to work load and deal flow priorities, we cannot seriously pursue it"

Figure 2: The Deal Flow: The firm screened 3631 deals, identified around 766 as interesting enough to evaluate, evaluated around 700 of those, and made 35 investments. We focus on the 66 deals that appeared interesting during screening but were rejected for capacity-related reasons. The composition of the 35 deals chosen for investment would probably have changed had the 66 deals been evaluated: A small number of the additional 66 deals would have replaced the worst of the best 35 deals selected from a sample of only 700. We ask: How much more valuable would a portfolio of the best 35 deals have been had they been drawn from the full population of 766 interesting deals instead of from the 700 deals that the venture-capital firm had the capacity to evaluate?

Screened Deals

	Not Eval.	Evaluated
Not Interest.	2865	0
Interesting	66	700

3631 → | → 35

Accepted Deals
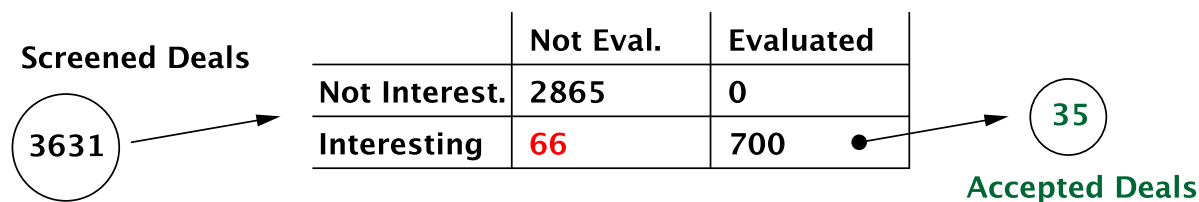
rejections. This represents an average of 2% of deals arriving for screening, varying on an annual basis from 1% to 4%. The available data does not permit us to determine whether these capacity-rejected deals were as attractive as those evaluated. The log comments establish, however, that the venture capitalists would have been likely to evaluate these deals had they not been limited by scarce capacity. Discussions within the company suggested that some of the deals rejected for capacity reasons were later observed to have a high value, causing the venture capitalists to question whether more capacity should be made available to permit the evaluation of more deals. We note in passing the statement by Tyebjee and Bruno (1984a, p. 207) that "good deals are more scarce than capital." Adding the 66 capacity-rejected deals would have resulted in a 9% increase in the total number of deals evaluated. The deal counts are summarized in Figure 2.

Figure 3: Selected Deals. This simulated example shows the top deals out of a sample of 700 or 766 drawn from the venture-capital-backed IPO data that we analyze in Section 5.1. The vertical lines depict the decision thresholds for the top 35 deals with (dashed) and without (solid) the extra 66 deals.
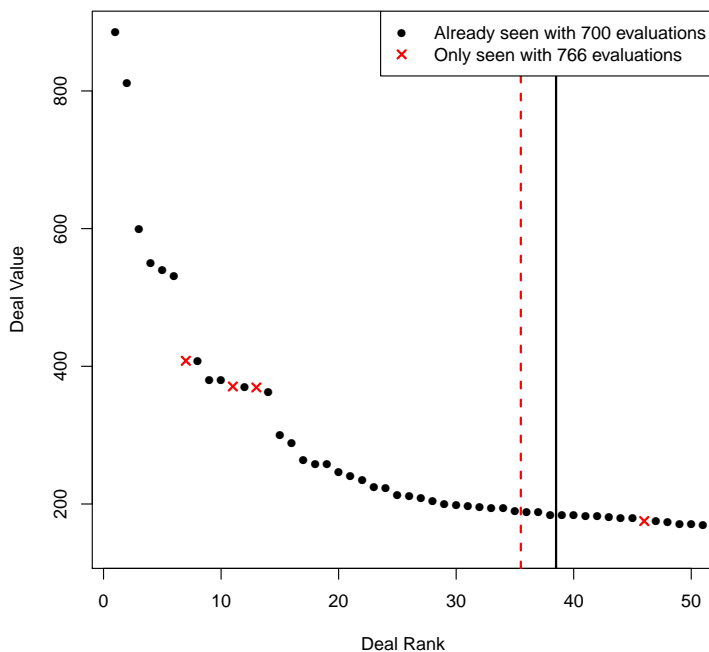

Figure 3 illustrates how increased sampling affects the composition of a portfolio made up of the top 35 deals. The solid points depict 700 deals that were first evaluated, and the X's depict the 66 extra deals evaluated given sufficient capacity to evaluate all deals classified as interesting during screening. The vertical lines depict the selection threshold for the top 35 deals with and without the 66 extra deals. Evaluating the additional 66 deals results in 3 deals being replaced by better deals. In this example, increasing the sample size from 700 to 766 deals does not change the value of the top six deals. The seventh deal with a sample of 766 deals has a value that is more than double the deal that it replaced. The deals ranked in eleventh and thirteenth place for 766 deals have a value that is more than 1.5 times the value of the deals that they replaced. The expected $\left(\frac{766}{700}\right)^{0.4} - 1 = 3.7\%$ increase in portfolio value from increasing the sample from 700 to 766 deals comes from a small number of the 35 deals radically increasing in value.

What is the expected value of increasing sampling by this magnitude if the 66 deals come from the same distribution as the original 700? How would this answer change if the capacity-rejected deals,

although interesting, were from a distribution with a lower expected value than the original deals? In Section 4 we develop the necessary theoretical tools, then return to these questions in Section 5.

4. Valuing Capacity

The goal of this section is to survey theoretical tools that can be used to answer our main question: How much would it be worth to a venture capital firm to increase the number of deals it evaluates by 9%? In other words, how does the value of the 35-deal portfolio the venture capitalists build grow as the number of evaluations increases from 700 to 766? This question is complex, and the answer hinges on many different factors such as:

- How does the quality of the deals that were rejected for capacity-related reasons compare to the quality of the deals that were already being evaluated?

- How competent are the venture capitalists? Are they skilled enough to pick out the best opportunities out of a larger sample of deals?

- How do interactions between deals impact the total fund value? Will a larger sample of potential deals help the firm build a more diverse portfolio?

- What are the time dynamics of fund investment? How does the timing of a deal's arrival influence its profitability if funded?

In an attempt to build a tractable model for the value of capacity, we begin by making some strong simplifying assumptions.

Suppose that the values of the capacity-rejected deals are indistinguishable from (identically distributed to) the deals that are already being evaluated; that the venture capitalists are perfectly skilled (can always recognize the best available deal when evaluating it); and that the venture firm is only allowed to invest in one deal and wants to maximize its return from this single deal (allowing us to side-step portfolio management issues, see Section 4.3). Then, the profit a venture-capital firm can make from any given sample of deals is fully determined by the value of best deal in the sample.

These fairly strong simplifying assumptions let us home in on the main driver of venture-capital profits, namely exceptional investment opportunities. A venture-capital firm that evaluates more deals will have more chances to encounter such opportunities.

The value of search depends on the distribution of deal values. At one end, suppose that all deal values are uniformly spread on [0, 1]. In this case, there is very little upside for search: Once we find one deal with a value close to 1 there is not much reason to look further. Conversely, if deal values have the same distribution as lottery-ticket earnings, then the expected value of the best deal grows almost linearly with the size of our sample because we are hoping to find the grand prize ticket. Our goal is to develop the technical machinery that will let us move beyond such caricatured examples and enable us to understand the value of search for realistic deal distributions.

We begin our analysis in Section 4.1 by using insights from extreme-value theory to address the question: How does the expected value of the best deal in a sample grow with sample size? Under our simplifying assumptions, the answer to this question is a close proxy to the value of capacity. As we will see, the value of increased sampling will be entirely determined by the shape of the tail of the distribution of deal values. Its mean and variance have asymptotically no impact on the value of increased sampling.[1]

In Appendix A, we study the sensitivity of our results to modeling assumptions. We allow for errors in venture capitalist valuations, and for time dynamics that force the firm to make decisions with incomplete information. Neither of these factors will end up affecting the relative value of increased evaluations.

4.1. The Growth Rate of Sample Maxima

Let F be the distribution of returns on potential venture-capital investment opportunities. Under regularity assumptions, the relationship between sample size and the expected value of sample maxima drawn from F takes a surprisingly simple form as sample sizes get large. The required regularity

---

[1]Similar insights can be derived from the extreme-value theoretic model of new-product development proposed by Dahan and Mendelson (2001). Dahan and Mendelson provide closed-form expressions to the new-product search problem under asymptotic conditions. We aim to derive simple relations that capture the main extreme-value theoretic insights while being straightforward enough for practitioners to use. In particular, our formulas can be expressed as proportionality relations, and do not depend on an asymptotic scale parameter.

assumption is that F have a *Pareto-type tail*. Formally, this requirement can be written as

$$\lim_{t\to\infty} \frac{1 - F(xt)}{1 - F(t)} = x^{-1/\gamma} \text{ for all } x > 0, \tag{1}$$

where $\gamma > 0$ is a constant called the *tail index* of F (see Coles (2001) or de Haan and Ferreira (2006) for a review). Heuristically speaking, F has Pareto-type tail whenever its density decays at a fixed polynomial rate at infinity. One way to understand the meaning of the tail index γ is that $\mathbb{E}[X^\nu]$ is finite for all $\nu < 1/\gamma$ and infinite for $\nu > 1/\gamma$.

This Pareto assumption—surprising at first glance—follows naturally from mathematical characterization theorems and has often been used successfully in practice. As argued by Chavez-Demoulin and Embrechts (2010), the Pareto assumption is often the best available tool for analyzing the behavior of large observations outside of a rigid parametric framework.

Under the Pareto assumption, it follows from classical results going back to Fisher and Tippett (1928) and Gnedenko (1943) that if we increase a sample size by a factor α, then we should expect the value of the best deal in the sample to increase by a factor α^γ. Formally, whenever X_i are independently and identically distributed according to a distribution F satisfying (1) for some $0 < \gamma < 1$, then

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[\max\{X_1, ..., X_{[\alpha n]}\}\right]}{\mathbb{E}\left[\max\{X_1, ..., X_n\}\right]} = \alpha^\gamma. \tag{2}$$

The required condition is satisfied, for example, by any Student-t distribution with more than 1 degree of freedom or any Pareto distribution with a finite mean.

This simple formula provides a first answer to our question. Provided that all our simplifying assumptions hold, that the distribution F has Pareto-type tails with tail index $0 < \gamma < 1$, and that our sample sizes are large enough that working in the limit where $n \to \infty$ is legitimate, then increasing the number of deals evaluated by the venture capitalist by 9% should increase return on investment by a factor $(1 + 0.09)^\gamma$. In other words, this formula lets us reduce the problem of modeling the relationship between sample size and sample maxima to the task of fitting the tail index γ.

Our discussion so far has made fairly strong assumptions. In Section 4.2 we explore how to adjust the model results when the dropped deals are of lower quality than the original set of deals chosen

to be evaluated. In Appendix A, we continue the theoretical discussion, and show that our result is robust to imperfect venture-capitalist skill and constrained time dynamics. In Section 5, we present evidence that the Pareto assumption may hold in the venture-capital context, and argue that $\gamma \approx 0.4$ should be reasonable for our application.

4.2. Adjusting for a Deal-Quality Differential among Dropped Deals

In our development so far, we have assumed that the deals that were flagged as interesting but then dropped due to capacity constraints were of the same quality as those that were evaluated. Venture capitalists would presumably prioritize the deals that appeared most promising at screening, thus it is useful to consider the scenario in which the dropped deals, although designated as worthy of evaluation, would nevertheless be of lower quality than those originally chosen for evaluation.

To address this question we compare the distributions of capacity-rejected and evaluated deals in terms of 1) their tail index and 2) the probability density under their right tail. Tail indices tend to be stable across similar datasets, and the tail index of a distribution is in general quite robust to noise.[2] It is thus common to assume in statistical applications that related distributions have the same tail index (Coles 2001, Davison and Smith 1990). We follow the standard practice of assuming that the dropped deals come from a distribution with the same tail index as that of the deals that were evaluated.

The impact of drawing the additional deals from a distribution with lower right-tail density can then be quantified by downweighting the sample size by the percentage that the probability mass is lower. The best outcome from a set of 100 deals drawn from a distribution with the same tail index as the evaluated deals but with 25% less probability mass in the upper tail would have roughly the same expected value as the best outcome of 75 deals drawn from the same distribution as the evaluated deals. A quality differential between the dropped and evaluated deals thus expresses itself as an incremental modification to our model, and does not invalidate the main insights. If, for example, the firm estimated the right-tail probability mass to be 10% lower for the capacity-rejected deals than for the evaluated deals, the 9% increase in sampling would be downweighted by 10% before applying the

---

[2]Our propositions 1 and 2 show that the effective tail index of the selected deal is not affected by noisy evaluations or time dynamics.

model.

Calculating the change in probability mass in the upper tail to determine how much to downweight the sample size of the dropped deals to compensate for a given quality differential requires data collection over time. While awaiting such data, managers can carry out a sensitivity analysis using a rough intuition about how the dropped and initially chosen deals compare.

4.3. Limitations: Allowing for More than One Accepted Deal

Our development until now has assumed that the venture capitalist can only invest in one deal. In this section, we show that our results can be extended to a portfolio made up of the k best deals, where k is fixed and small relative to the sample size. Then, if we write $X_{1,n} \geq ... \geq X_{n,n}$ for the sorted values of the evaluated deals, the venture capitalist will earn $S_k^{(n)} = \sum_{i=1}^{k} X_{i,n}$ with noiseless evaluations. With noisy evaluations we get an analogous expression. Provided that k remains fixed as n goes to infinity, our main results still hold and the expected portfolio value grows asymptotically by a factor α^{γ} as we increase the number of deals evaluated by a factor of α. Heuristically, the reason for this is that $X_{1,n}$, ..., $X_{k,n}$ all scale at the same rate as n gets large, so $\mathbb{E}[S_k^{(n)}]/\mathbb{E}[X_{1,n}]$ converges to a finite limit when $n \to \infty$; see Appendix C for details.

In a nutshell, the classic result—that increasing the sample size by a factor of α is expected to increase the value of the maximum value by a factor of α^{γ}—also applies to the deals in second, third, and so on to kth position. Thus, an increase in sample size will also cause the portfolio that results from combining these k deals to increase by the same factor.

The picture becomes more complicated if the firm can also use the number k of accepted deals as a decision parameter, or if the accepted deals can interact. Extending our results to a context where the process of deal selection has an interplay with portfolio management issues is an interesting topic for further research.

5. Applying the Model to the Capacity Decision

The model we advocated in the previous section can be summarized as follows

If the number of deals evaluated increases by a factor α, $\hspace{3cm}$ (3)

Then the expected value of the selected portfolio rises by a factor α^γ,

where $\gamma > 0$ is the tail index of the distribution of available investment opportunities as defined in (1). In this section we apply our model to the decision by the firm of how many deals to evaluate. We first use IPO return data to gain intuition about the tail weight, then compare the resulting increase in fund value to a rough estimate of the cost of allocating more capacity to deal evaluation.

5.1. Fitting the Tail Index

Fitting the distribution F of potential venture returns from empirical data is challenging. Venture capitalists typically realize profits either through IPOs or mergers and acquisitions (M&A), the latter being more common (Giot and Schwienbacher 2007, NVCA 2010). M&A data is usually kept private, however, and the original valuation of investments that result in IPOs is often not known (Kaplan and Schoar 2005). Sandberg (1987) noted, for example, that ventures backed by venture capital survive at a higher rate than ventures lacking such backing. That venture funding impacts venture return creates an endogeneity problem.

Noting these difficulties, we observe that venture-capital-backed IPOs are part of the set of deals evaluated by venture capitalists. Information on IPOs is publicly available. For our analysis, we used the Thomson Financial SDC Platinum New Issues database, which contains information for 2,037 venture-capital-backed IPOs in the United States between 1995 and 2011. IPO data provides a useful proxy that can be used to gain insight about the true distribution of venture capital opportunities.

Under the strong assumption that initial venture-capital investments are roughly independent from IPO proceeds, the tail index of the distribution of gross IPO proceeds should be indicative of the tail index of potential venture-capital returns, with a further caveat that the data analyzed is not corrected for market fluctuations or for inflation.

Figure 4: Fitting the Tail Index.

(a) RBM and Hill estimates for the tail index γ.

(b) Mean residual life plot.
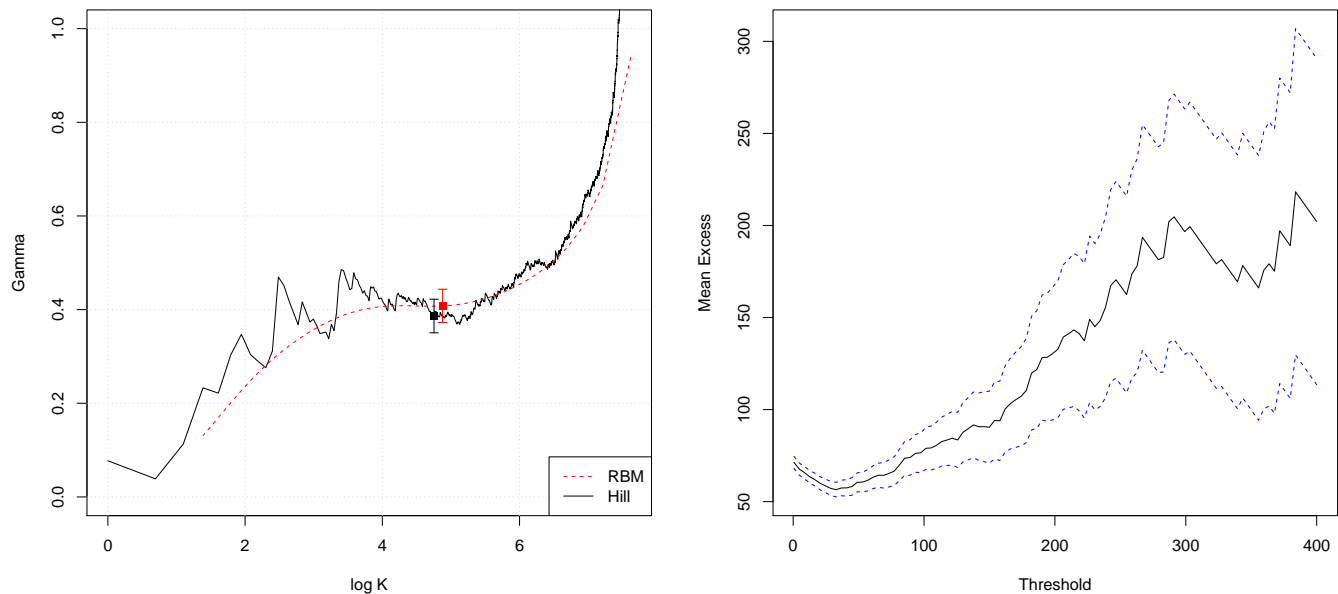
Back-fitting a slight modification of model (2) to randomly selected subsets of the IPO data gives rise to the Random Block Maxima (RBM) estimator (Wager 2014). A closely related and widely used estimator is the Hill estimator (Hill 1975). Both the RBM and Hill estimators require choosing a tuning parameter k. Low values of k lead to high variance and high values of k are often biased, making the choice of tuning parameter crucial. For the RBM estimator, we choose k following the rule recommended by Wager (2014); with the Hill estimator we use a diagnostic for bias developed by Guillou and Hall (2001).

A third estimate of γ is made by fitting a generalized Pareto tail to the data by maximum likelihood, as described in Coles (2001). This method requires fixing a threshold u that cuts the tail from the body of the distribution. Coles (2001) does not provide an automatic method for doing this, but rather suggests examining a mean residual life plot and fixing u such that the mean residual plot becomes linear to the right of u.

The RBM estimator gives a 95% confidence interval of $\gamma = 0.41 \pm 0.07$, while the Hill estimator gives $\gamma = 0.39 \pm 0.07$. Plots for the RBM and the Hill estimators as a function of the tuning parameter

k are given in Figure 4a. The plotted error bars at the automatically selected values of k are one standard deviation wide. (These error bars do not account for adaptive selection of k.) The stability of both the RBM and Hill estimators for $\log k$ between 4 and 6 is encouraging, and suggests that the Pareto assumption (1) in fact holds.

For maximum likelihood estimation, we chose a threshold of $u = 160$ using the mean residual life plot in Figure 4b. This leaves us with 155 data points above the threshold, and yields a 95% confidence interval of $\gamma = 0.47 \pm 0.24$. The mean residual life plot was generated using the R package ismev (Stephenson 2011). Although the plot is defined up to $u = 800$, we only display the first half, as the second half is too noisy to be meaningful.

These results are consistent with the distribution F of potential returns on ventures being in the domain of attraction $\mathcal{D}(G_\gamma)$ of an extreme-value distribution with positive tail index $\gamma \approx 0.4$.[3] The relatively heavy tail of IPOs suggests that venture capitalists should consider the possibility that the distribution of deal values has a positive tail index.

5.2. Trade-off Analysis

How much capacity would need to be made available for the firm to increase the number of deals evaluated? We can make a rough estimate of the amount of venture-capitalist work represented in evaluating a deal. After the initial year, team members processed on average 78.7 proposals each year. The mean team size was 4.6. Robinson (1987) observed that venture capitalists spend about 33% of their time on evaluation. Assuming that the venture capitalists worked 230 days per year, devoted a third of their time to evaluation, and evaluated 20% of the processed deals, the number of days required to evaluate a deal averaged just under 5. Thus, to evaluate one additional deal would require the allocation of around 1 week of venture-capitalist capacity assuming that the additional deals to be evaluated were reasonably spread out over the period.

To get a feeling for how much such a capacity increase might cost, we can consider the median salaries for venture capitalists. According to Salary.com, the median expected salary for a junior venture capitalist in the U.S. (venture capitalist I) in early 2012 was \$164,228, rising to \$276,801 for a

---

[3] One way to understand this number is that, for $\gamma > 0$, if a distribution G has a tail index γ, then G has ν^{th} moments only for $\nu < \gamma^{-1}$. A tail index of 0.4 implies that F has a finite variance but no third moment.

Table 3: The expected percentage increase in fund value from increased evaluations

| | TAIL INDEX | | | | |
INCREASE	0.2	0.3	0.4	0.5	0.6
5%	1%	1.5%	2%	2.5%	3.0%
10%	1.9%	2.9%	3.9%	4.9%	5.9%
15%	2.8%	4.3%	5.7%	7.2%	8.7%
20%	3.7%	5.6%	7.6%	9.5%	11.6%

Table 4: The expected increase in fund value (in $mm) from increased evaluations given a base fund value of $151,000,000

| | TAIL INDEX | | | | |
INCREASE	0.2	0.3	0.4	0.5	0.6
5%	$1.5	$2.2	$3.0	$3.7	$4.5
10%	$2.9	$4.4	$5.9	$7.4	$8.9
15%	$4.3	$6.5	$8.7	$10.9	$13.2
20%	$5.6	$8.5	$11.4	$14.4	$17.5

senior venture capitalist (venture capitalist III) (salary.com 2012). Thus, the salary cost of evaluating an extra deal can be expected to be between $3000 and $6000. If the venture-capital firm increased the deal evaluation rate by 10% from 20% to 22%, this would imply an increase in costs on the order of magnitude of $200,000 to $400,000, depending on whether the additional capacity came from junior or senior venture capitalists. The firm may also be able to reduce the cost of increased evaluations through a reallocation of resources from other activities, through allocating some evaluation tasks to less expensive employees, or through use of decision-support tools (Zacharakis and Meyer 2000). Our estimates can thus be considered as an upper bound on the cost of creating the capacity necessary to do the extra evaluations.

We apply model (3) to estimate the expected increase in the value of the fund under tail weights ranging from 0.2 to 0.6 when evaluations are increased by 5% to 20% as shown in Table 3. In Table 4 we apply the Table 3 percent increases to the average U.S. venture-capital fund size, estimated by the National Venture Capital Association for the U.S. at $151,000,000 in 2009 (NVCA 2010).

A 9% increase in evaluations at a tail weight of 0.4 corresponds to an increase of 3.5%—$5,300,000 for a fund of average size—if the additional deals come from the same distribution as the original deals. We estimated that the cost of evaluating the extra deals using a senior venture capitalist would

be less than $500,000, thus the expected increase substantially exceeds the expected capacity cost.

Now let's consider the possibility that the extra deals are drawn from a distribution with the same tail index as that of the original deals, but a probability mass under the tail that is 44% lower. In this case the impact of a 9% increase in evaluated deals only has the impact of a $9\% \times (1-44\%) = 5\%$ increase. At a tail index of 0.4, the value of a 5% increase can be seen from Table 4 to be $3,000,000, still substantially greater than the cost of capacity.

Until now we have only considered deals that were explicitly minuted by venture capitalists as worth evaluating in estimating the value of increased sampling. This sensitivity analysis raises an interesting issue: Under positive tail weight, should the firm screen less aggressively? Suppose that the firm were to reduce the threshold for "interesting" so that about 30% of deals would pass screening. Suppose that these extra deals would have a probability mass under the tail of only 25% of that of the original deals, so that the 50% increase in sampling would be downweighted by 75%. At a 0.4 tail weight this $50\% \times 25\% = 12.5\%$ increase would result in a 4.8% increase in portfolio value, again generating enough return to pay for an increase in capacity even though the additional deals were clearly less valuable than the original set selected.

Our analysis has ignored other benefits that could emerge from reducing utilization in the venture-capital operations. A modest reduction in utilization would be expected to substantially reduce lead times. Rejections went both ways: The firm rejected deals, but was also rejected by entrepreneurs. In 283 cases (representing almost half of the deals evaluated), the entrepreneur withdrew his or her proposal before the fund had a chance to make an investment decision. The venture-capital firm contacted the entrepreneur to request further information and received no response in 177 cases. Another venture-capital firm was chosen, or the firm was not invited to join a syndicate, in 73 cases. The entrepreneur decided against venture-capital financing in 33 cases. Research suggests that customers are sensitive to lead time (e.g., Allon et al. 2011). To the degree that shorter lead times would have reduced these rejections by entrepreneurs, the gain in fund value from implementing a chase capacity strategy might be greater than predicted by our model.

5.3. Monte Carlo Experiments

The model presented in (3) is technically only valid in the limit $n \to \infty$, where n is the number of evaluated deals. The firm chose to invest in 35 out of approximately 700 evaluated deals. To what degree do limiting results for $n \to \infty$ apply to this situation? Such issues are a recurrent concern in applications of extreme-value theory, and suggest that we should interpret the results obtained using the equation (3) model with caution. The magnitude of the impact from increasing the number of evaluations that emerges from our analysis is substantial enough, however, that the numbers in Table 4 may be reasonably taken as useful guidance.

To test for small-sample issues, we ran a Monte Carlo study on the IPO dataset used to fit γ. We repeatedly drew 700 and 766 random deals from our dataset of 2,037 venture-backed IPOs, and compared the sum of the top 35 deals from both subsamples. According to our formula, the marginal relative difference between the expectations should be $\left(\frac{766}{700}\right)^{0.4} - 1 = 3.7\%$. In the Monte Carlo simulation with 1,000,000 replicates, this number came out to be 3.6%. Thus, at least in the context of the IPO dataset, our sample sizes appear to have been large enough to justify the use of extreme-value-theoretic methods.

We also tested the performance of our model with evaluation noise. To emulate sub-optimal VC decision processes, we multiplied the values of the available deals with mean-1 log-normal noise with shape parameter $\sigma = 1/3$ and used the result as noisy estimates of the deal values. We then picked out the 35 deals with the highest estimated values, rather than selecting the deals with the highest actual values. With this noisy decision process the marginal value of evaluating 766 versus 700 deals drops to 3.5%. Thus, when we introduce errors into the VC decision making process, our model is not quite as accurate as in the error-free setup but can still give us a fairly good intuition for the value of additional sampling.

6. Applying the Model to Personnel Selection

In this section we explore application of our model to the question of how many candidates to evaluate in making a hiring decision. Our analysis is motivated by the following situation faced by a large international organization. A division received 500 applications for a post to be filled, and the

Human Resources department did a rough first screening to reduce the applicant pool by half. The director of the department for which the hire was being made observed that Human Resources had excluded two candidates that he would have ranked as worth a closer look. Concerned that strong candidates had been overlooked, he devoted over 20 hours of time to reviewing the dossiers than had been rejected by Human Resources. We used the model to estimate the expected payoff from the investment of the director's time in terms of improvement in the hiring outcome.

Let's define a hiring outcome as the ratio between the performance level of a candidate and a performance level that meets but does not exceed the expectations of the job. The mean of the distribution of hiring outcomes is assumed to lie at 1, so that an average applicant is just able to do the job. For the post in question, the intuition was that a really great hire (the best of 500) could be expected to perform around ten times as well as a minimally qualified candidate. The best of 50 candidates might be expected to perform about five times as well.

We can reverse-engineer our model to turn this intuition—that the best candidate out of a pool of 500 should be twice as good as the best candidate out of a pool of 50—into an estimate of the tail index. A 10-fold increase in evaluations should increase the expected value of the maximum by a factor 10^γ. Transforming these two points into a curve that links sample size and the expected value of the maximum requires that we solve

$$10^\gamma = 2, \text{ which has a solution } \gamma = \frac{\log 2}{\log 10} \approx 0.3.$$

Applying the model assuming a tail index of 0.3, we estimate that doubling the number of candidates evaluated should increase the performance of the top candidate by $2^{0.3} - 1 = 23\%$. Let's first explore a 50% reduction in the sample size done without any consideration of quality, corresponding to decreasing the applicant pool from 500 to 250. If the best of 500 candidates is expected to perform ten times better than a minimally qualified hire, the model predicts that a 50% decrease would result in a decline in the level of the best candidate relative to the mean of $2^{0.3} - 1 = 23\%$, that is, at a level of 7.7 rather than ten. Such a difference may well justify 20 hours of additional candidate screening.

Now let's imagine a scenario in which nine of ten applicants rejected by the Human Resources department were in fact almost certain to not be a top candidate. One in ten, however, was good

enough to pursue further. Thus, the tail weight of the distribution of rejected applicants would remain 0.3, but the probability mass under the tail would be reduced by 90%. Applying the model extension developed in Section 4.2 we estimate that doubling the applicants considered by the director of the hiring department from 250 to 500 from a distribution with 90% less probability mass under the tail corresponds to an increase of $250 \times 10\% = 25$ or a 10% increase in applicants from the original distribution. A 10% increase in applicants is valued by the model at a $1.1^{0.3} - 1 = 3\%$ increase in hiring outcome, which does not appear likely to warrant the investment of 20 hours of additional screening. This simple example shows how the model can facilitate incorporating tail weight in deciding how much time to allocate to sampling, and how even a rough intuitive estimate of tail weight can be useful in quantifying the expected benefit of increased capacity.

6.1. Finite-n Analysis

The relationship between the sample size and the expected value of the best outcome holds as n goes to infinity. The personnel-selection example can be enriched to give a sense of the impact of finite n on model results. Suppose that an oracle has provided information about the underlying performance distributions for two positions. Both have a mean of 1 and coefficient of variation of 0.2. The thin-tailed distribution is $\mathcal{N}(1, 0.2^2)$. The other follows a generalized-extreme-value distribution (GEV, also known as the Fréchet) with a tail index $\gamma = 0.3$, mean of 1, and scale and location parameters scaled to obtain the 0.2 coefficient of variation. The two densities are shown in Figure 5. Knowing the distributions allows us to sample and explore the behavior of the sample maxima. For the two distributions, we calculated the average of sample maxima over 100,000 replications, doubling the sample size from 25 to 3200. The results are shown in Table 5. The first column shows the sample size. The average of the sample maxima for the normal and GEV distributions are given in the second and third columns.

For the normal outcome distribution, increasing the sample size from 25 to 50 yields a 4% payoff. A second doubling to 100 increases the expected best outcome by 3%, and a third to 200 increases it by an additional 3%. An increase from 1600 to 3200 yields a 2% increase. For the heavy-tailed distribution, increasing the sample size from 25 to 50 or from 50 to 100 yields a 14% increase, and a third doubling to 200 yields an increase of 16%. Doubling the sample size from 1600 to 3200 yields

Figure 5: The figure compares normal and GEV densities that describe the distribution of job-related capability for a given position. Both distributions have mean 1 (representing a capability level that exactly matches minimum job requirements) and coefficient of variation 0.2. The GEV has a tail index of 0.3. For applicant pools whose capability distribution is normal, an outcome above 1.7 is not expected even with a sample size of 3000 or more. For applicant pools with a GEV capability distribution, the plot shows how the GEV density hovers above the horizontal axis well beyond a capability level of 1.7. Our simulation shows that relative performance levels of 3 or more are expected by the time we reach a sample size of 1000.
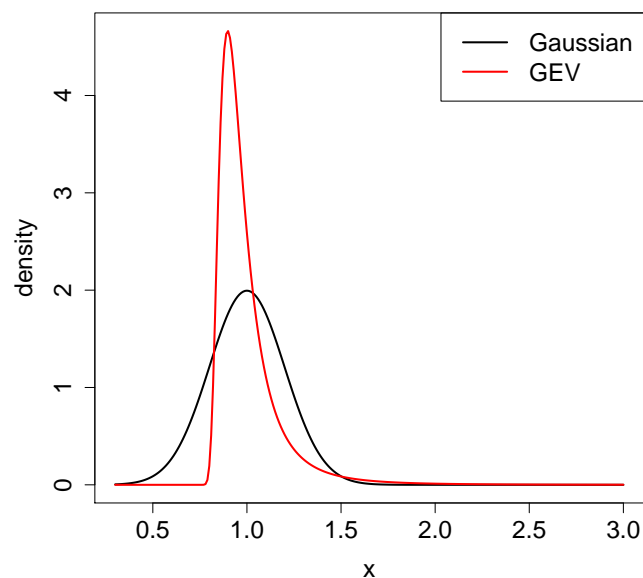

Table 5: The maximum value from a sample of size N, averaged over 100,000 replications.

N	Normal	GEV
25	1.39	1.58
50	1.45	1.80
100	1.50	2.06
200	1.55	2.39
400	1.59	2.79
800	1.64	3.29
1600	1.68	3.89
3200	1.71	4.65

a 20% increase. As sample sizes grow to infinity, the expected gain from doubling the sample size converges to 23%, which is the gain predicted by our formula (3). The full distribution information thus gives us a better understanding of the relevance of the model for finite samples. These results combine with the Monte Carlo simulation described in the previous section to suggest that the model

results can provide useful managerial insight even for finite sample sizes.

7. Conclusion

This research project emerged from our observation that a venture-capital firm rejected otherwise interesting deals because of capacity limitations, consistent with use of a lean heuristic to encourage efficiency and productivity gains through the use of exploratory stress. Firms must choose when to emphasize elimination of uncertainty, and when to follow a chase capacity strategy, investing in a capacity buffer to gain value from that uncertainty.

We have shown that the value of search depends crucially on extremes in the distribution of outcomes. The cost of capacity is highly visible, whereas the value of increased search is obscure. In cases where uncertainty is driven by tail weight rather than by a second moment, the capacity-allocation decision becomes less intuitive. Applying tools from extreme-value theory reduces the risk of underinvesting in capacity. Data on returns from venture-capital-backed IPOs provided a general idea of the tail index of the deal-value distribution. Model results suggest that the value of increasing the number of deals evaluated appears to substantially exceed the cost of the extra capacity required.

The value of a venture-capital portfolio depends on the value of the top few deals available for evaluation. Such an extremes-driven setup goes counter to traditional operations-management thinking. Our question, "How does the expected sum of the best 35 deals from a pool of 700 compare to the expected sum of the best 35 deals from a pool of 766?" can only be systematically addressed using extreme-value theory. That the value of the portfolio grows at the same rate as the value of the best few deals as we deepen the search opens the door to the use of mathematical tools that inform the capacity-allocation decision. Recognizing that portfolio value depends on extremes rather than averages may give additional insights into process improvements. A screening process oriented toward averages rather than extremes risks rejecting deals with tail potential that do not look promising according to normal heuristics.

Our model development was inspired by the capacity decision faced by the venture-capital firm. We then used a simple example to illustrate how these results can be generalized to personnel selection. How extensively should a firm cast its net when seeking to fill a position? The performance-value

distribution for some posts will be thin tailed, where a superb candidate will do better than an adequate candidate, but the increase in value will be modest. For other posts, the performance-value distribution will have a heavy right tail, so that a superb candidate will create value that is far beyond that of an adequate candidate. Although it is not obvious how to fit the tail index of the outcome distribution, we have illustrated how even a relatively qualitative discussion of the concept can aid in thinking through the sampling strategy. The same logic can be applied to other search operations such as root-cause analysis. Plant managers say that they recognize the value of identifying root causes, but easily scrimp on allocating capacity to the search process (MacDuffie 1997). Again, a rough estimate of the right-tail weight can help to motivate appropriate investment in root-cause search activities.

At a higher level, 3M's allocation of 15% of employee capacity to innovation has been generally considered to be one of the reasons for the company's high percentage of sales from highly innovative products (`solutions.3m.com`). In recent years, top management has reduced the capacity available for pure innovation, and the percent of sales from highly innovative products has decreased (Hindo 2007). At the same time, 3M has put into place a system to incorporate lead users into the design process (Lilien et al. 2002). Our model provides a way to structure the decision about how much capacity to invest in innovation. Initiatives like including lead users in product design can be evaluated in terms of their ability to increase right-tail weight. The tail weight can then be used to determine where pure-innovation capacity is likely to pay off.

Applying the model requires that we convince decision makers to incorporate richer information about distributions in their models. This can be done: Supply chain management research was galvanized in the mid-1990s by the realization that dispersion in forecasters' estimates could provide information about the standard deviation of the demand distribution (e.g., Fisher and Raman 1996). The ability to easily estimate the standard deviation of demand dramatically increased willingness to include variation in decision making, for practitioners as well as academics. Terwiesch and Ulrich (2009) proposed a similar approach for estimating the central moments of the outcome distribution in the innovation context. A next step is to consider how to use upside potential rather than just central moments in screening. Fleming (2007, p. 74) indicated that such a step is worth consideration:

"by understanding the histogram of inventive creativity and managing the factors that shape that distribution, a firm can greatly improve its capacity to innovate..."

Singh and Fleming (2010) expressed concern that incorporation of tail weight in the distribution of innovation outcomes, although essential to sound decision making, risked increasing the complexity of research (see also Erat and Kavadias 2008). It is generally known, however, that under heavy-tail conditions, most questions of managerial interest can be answered by only referring to the tail index of the distribution. Our model builds on that insight. Through formulating the answer to the question "What is the marginal benefit of sampling more deals?" solely in terms of the tail index, counter-intuitive insights from extreme-value theory become accessible to the world of practice.

References

References

Allon, G., Federgruen, A., Pierson, M., 2011. How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. Manufacturing & Service Operations Management 13 (4), 489–507.

Chao, R., Kavadias, S., Gaimon, C., 2009. Revenue driven resource allocation: Funding authority, incentives, and new product development portfolio management. Management Science 55 (9), 1556–1569.

Chavez-Demoulin, V., Embrechts, P., 2010. Revisiting the edge, ten years on. Communications in Statistics—Theory and Methods 39 (8-9), 1674–1688.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer Verlag.

Dahan, E., Mendelson, H., 2001. An extreme-value model of concept testing. Management Science 47 (1), 102–116.

Davison, A. C., Smith, R. L., 1990. Models for exceedances over high thresholds. Journal of the Royal Statistical Society. Series B (Methodological), 393–442.

de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer Verlag.

Erat, S., Kavadias, S., 2008. Sequential testing of product designs: Implications for learning. Management Science 54 (5), 956–968.

Fisher, M., Raman, A., 1996. Reducing the cost of demand uncertainty through accurate response to early sales. Operations Research 44 (1), 87–99.

Fisher, R., Tippett, L., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society 24 (2), 180–190.

Fleming, L., 2007. Breakthroughs and the "long tail" of innovation. Sloan Management Review 49 (1), 69–74.

Fried, V. H., Hisrich, R. D., 1994. Toward a model of venture capital investment decision making. Financial Management 23 (3), 28–37.

Gifford, S., 1997. Limited attention and the role of the venture capitalist. Journal of Business Venturing 12 (6), 459–482.

Giot, P., Schwienbacher, A., 2007. IPOs, trade sales and liquidations: Modelling venture capital exits using survival analysis. Journal of Banking & Finance 31 (3), 679–702.

Girotra, K., Terwiesch, C., Ulrich, K., 2010. Idea generation and the quality of the best idea. Management Science 56 (4), 591–605.

Gnedenko, B., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. The Annals of Mathematics 44 (3), 423–453.

Goodale, J., Verma, R., Pullman, M., 2003. A market utility-based model for capacity scheduling in mass services. Production and Operations Management 12 (2), 165–185.

Gorman, M., Sahlman, W. A., 1989. What do venture capitalists do? Journal of Business Venturing 4 (4), 231–248.

Guillou, A., Hall, P., 2001. A diagnostic for selecting the threshold in extreme value analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (2), 293–305.

Hall, J., Hofer, C. W., 1993. Venture capitalists' decision criteria in new venture evaluation. Journal of Business Venturing 8 (1), 25–42.

Hayes, R., Wheelwright, S., 1984. Restoring our Competitive Edge: Competing through Manufacturing. Wiley.

Hill, B., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3 (5), 1163–1174.

Hindo, B., June 11 2007. At 3M, a struggle between efficiency and creativity. Business Week, 8–14.

Hopp, W., Spearman, M., 2001. Factory Physics: Foundations of Manufacturing Management. McGraw-Hill/Irwin.

Jack, E., Raturi, A., 2002. Sources of volume flexibility and their impact on performance. Journal of Operations Management 20 (5), 519–548.

Jessen, A., Mikosch, T., 2006. Regularly varying functions. Publications de l'Institut Mathématique, Nouvelle Série (Beograd) 80 (94), 171–192.

Kanniainen, V., Keuschnigg, C., 2003. The optimal portfolio of start-up firms in venture capital finance. Journal of Corporate Finance 9 (5), 521–534.

Kaplan, S., Schoar, A., 2005. Private equity performance: Returns, persistence, and capital flows. The Journal of Finance 60 (4), 1791–1823.

Kellogg, D., Nie, W., 1995. A framework for strategic service management. Journal of Operations Management 13 (4), 323–337.

Kennedy, D., Kertz, R., 1991. The asymptotic behavior of the reward sequence in the optimal stopping of iid random variables. The Annals of Probability 19 (1), 329–341.

Koste, L., Malhotra, M., Sharma, S., 2004. Measuring dimensions of manufacturing flexibility. Journal of Operations Management 22 (2), 171–196.

Lilien, G. L., Morrison, P. D., Searls, K., Sonnack, M., Von Hippel, E., 2002. Performance assessment of the lead user idea-generation process for new product development. Management science 48 (8), 1042–1059.

Loch, C., Kavadias, S., 2002. Dynamic portfolio selection of NPD programs using marginal returns. Management Science 48 (10), 1227–1241.

MacDuffie, J., 1997. The road to "root cause": Shop-floor problem-solving at three auto assembly plants. Management Science 43 (4), 479–502.

Macmillan, I. C., Siegel, R., Narasimha, P. N. S., 1985. Criteria used by venture capitalists to evaluate new venture proposals. Journal of Business Venturing 1 (1), 119–128.

Maier, J., Walker, I., 1987. The role of venture capital in financing small business. Journal of Business Venturing 2 (3), 207–214.

Monden, Y., 2011. Toyota Production System: An Integrated Approach to Just-in-Time, 4th Edition (Kindle). Institute of Industrial Engineers, CRC Press—Taylor & Francis.

NVCA, 2010. National Venture Capital Association Press Center FAQ. Arlington, VA, verified January 21, 2011.
URL `http://www.nvca.org/index.php?Itemid=147&id=119&option=com_content&view=article`

Riquelme, H., Rickards, T., 1992. Hybrid conjoint analysis: An estimation probe in new venture decisions. Journal of Business Venturing 7 (6), 505–518.

Roberts, E., 1991. Entrepreneurs in High Technology: Lessons from MIT and Beyond. Oxford University Press, USA.

Robinson, R. B., 1987. Emerging strategies in the venture capital industry. Journal of Business Venturing 2 (1), 53–77.

Sahlman, W., 1990. The structure and governance of venture-capital organizations. Journal of Financial Economics 27 (2), 473–521.

salary.com, 2012. Salary Wizard. Verified March 2, 2012.
URL `http://http://www1.salary.com/Venture-Capitalist-I-Salary.html`

Sandberg, C., 1987. Improving new venture performance: The role of strategy, industry structure, and the entrepreneur. Journal of Business Venturing 2 (1), 5–28.

Sapienza, H., Gupta, A., 1994. Impact of agency risks and task uncertainty on venture capitalist-CEO interaction. Academy of Management Journal 37 (6), 1618–1632.

Shah, R., Ward, P. T., 2003. Lean manufacturing: Context, practice bundles, and performance. Journal of Operations Management 21, 129–149.

Shepherd, D., Armstrong, M., Levesque, M., 2005. Allocation of attention within venture capital firms. European Journal of Operational Research 163 (2), 545–564.

Silverberg, G., Verspagen, B., 2007. The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. Journal of Econometrics 139 (2), 318–339.

Singh, J., Fleming, L., 2010. Lone inventors as sources of breakthroughs: Myth or reality? Management Science 56 (1), 41–56.

Stephenson, M., 2011. ismev: An Introduction to Statistical Modeling of Extreme Values. R package version 1.36, verified March 12, 2012.
URL `http://CRAN.R-project.org/package=ismev`

Suri, R., 1998. Quick Response Manufacturing. Productivity Press, Portland, OR.

Suri, R., de Treville, S., 1986. Getting from Just-in-Case to Just-in-Time: Insights from a simple model. Journal of Operations Management 6 (3-4), 295–304.

Terwiesch, C., Loch, C., 2004. Collaborative prototyping and the pricing of custom-designed products. Management Science 50 (2), 145–158.

Terwiesch, C., Ulrich, K., 2009. Innovation Tournaments: Creating and Selecting Exceptional Opportunities. Harvard Business School Press, Boston, MA.

Terwiesch, C., Xu, Y., 2008. Innovation contests, open innovation, and multiagent problem solving. Management Science 54 (9), 1529–1543.

Tyebjee, T., Bruno, A., 1984a. Venture capital: Investor and investee perspectives. Technovation 2 (3), 185–208.

Tyebjee, T. T., Bruno, A. V., 1984b. A model of venture capitalist investment activity. Management Science 30 (9), 1051.

Van Mieghem, J., 2003. Commissioned paper: Capacity management, investment, and hedging: Review and recent developments. Manufacturing & Service Operations Management 5 (4), 269–302.

Wager, S., 2014. Subsampling extremes: From block maxima to smooth tail estimation. Journal of Multivariate Analysis 130, 335–353.

Weitzman, M., 1979. Optimal search for the best alternative. Econometrica: Journal of the Econometric Society, 641–654.

Womack, J. P., Jones, D. T., Roos, D., 1990. The Machine that Changed the World. Harper Collins Publishers, New York.

Zacharakis, A. L., Meyer, G. D., 2000. The potential of actuarial decision models: Can they improve the venture capital investment decision? Journal of Business Venturing 15 (4), 323–346.

Acknowledgements

Appendix A Robustness Analysis

We established the formula (2) under quite heavy-handed simplifying assumptions. Perhaps the least reasonable of our assumptions is that venture capitalists can consistently pick out the best deals out of a pool of evaluated deals. There is considerable evidence that venture capitalists regularly make suboptimal decisions (Zacharakis and Meyer 2000); moreover, venture capitalists are often bound by tight time constraints that force them to make investment decisions on a rolling basis without knowledge of what opportunities may be available in the future.

In Section A.1 we loosen the assumption of venture capitalist infallibility by adding in multiplicative noise in evaluations. Thus, the deal selected by the venture capitalist from the sample $\{X_1, ..., X_n\}$ may not coincide with the sample maximum. We denote the value of the selected deal by $\text{Select}(\{X_1, ..., X_n\})$. In Section A.2, we relax the assumption that venture capitalists have simultaneous access to the full set of deals to be evaluated.

A.1 Valuation Errors

A simple way to model imperfect venture capitalist skill is by using multiplicative valuation errors: Suppose that if a deal actually has value X_i, then the venture capitalist values it at $V_i = E_i \cdot X_i$, where $E_i \sim G$ is a noise term that is independent of X_i. Under this multiplicative noise model,

$$\text{Select}(\{X_1, ..., X_n\}) = X_{i^*}, \text{ where } i^* = \text{argmax}_{1 \leq i \leq n}\{E_i \cdot X_i\}. \tag{4}$$

We chose to model venture capitalist errors as multiplicative rather than additive ($V_i = X_i + E_i$), because it seems plausible that venture capitalists make larger valuation errors on more extreme deals.

Clearly, such valuation errors will reduce the expected of value deals selected by a venture capitalist: $\mathbb{E}\left[\text{Select}(\{X_1, ..., X_n\})\right] < \mathbb{E}\left[\max\{X_1, ..., X_n\}\right]$. However, as we show below, the cost of such valuation errors converges to a multiplicative factor that does not depend on sample size. This multiplicative factor cancels out when we take a ratio; thus valuation errors have asymptotically no impact on the relative value of increased evaluations and the formula (5) holds.

Proposition 1. *Let $V_i = E_i \cdot X_i$, where $X_i \sim F$ and $E_i \sim G$ independently for distributions F and G whose support is bounded below by $B > 0$. Suppose that F has a tail index $0 < \gamma < 1$, and that there*

is a constant $\varepsilon > 0$ such that $\mathbb{E}\left[E_i^{1/\gamma+\varepsilon}\right] < \infty$. Under these conditions,

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[\mathrm{Select}(\{X_1, ..., X_{[\alpha n]}\})\right]}{\mathbb{E}\left[\mathrm{Select}(\{X_1, ..., X_n\})\right]} = \alpha^\gamma \tag{5}$$

holds for the selection function (4).

The main assumption made for this result is that the tail of the distribution G of the error terms decays faster than the tail of F: recall that X_i has no ν^{th} moments for $\nu > \gamma$, and so this main assumption is enforced by the technical condition that $\mathbb{E}\left[E_i^{1/\gamma+\varepsilon}\right] < \infty$ for some $\varepsilon > 0$. Heuristically, this assumption states that very large errors are smaller than very large X-values, and so the signal is not completely swamped by the noise.

The simple formula (5) shows that our main result (3) holds even in the face of evaluation error. The overall value of the portfolio decreases in venture capitalist error, but the relative increase in the value of the best deal with increased sampling remains a simple function of the tail index. Our capacity-valuation heuristic, namely that increasing the number of evaluated deals by 9% would increase return on investment by a factor 1.09^γ, thus appears to be robust to modeling assumptions.

A.2   Time Dynamics

Modeling a venture capitalist as simultaneously evaluating all received deals and choosing the best of them—as we have done so far—is not a realistic picture of venture-capital time dynamics. Deals arrive over time, and venture capitalists must often make a decision about a deal having arrived for evaluation without the luxury of waiting to see what potentially more interesting deals might arrive later. That the venture capitalist must make a decision without the full information that would come from having all deals reduces the expected value of the evaluation operation.

We consider the worst-case scenario for a venture capitalist trying to select the best deal out of a large sample: The venture capitalist is looking to invest in exactly one deal, is only allowed to see one deal at a time, and must make an investment decision on that deal before seeing the next deal. As shown below, such timing constraints have a similar impact on venture capitalist profits as valuation errors: They reduce the expected value of the selected deal by a multiplicative factor that—asymptotically—does not depend on sample size. Thus, our formula (3) still holds.

Proposition 2 (Kennedy and Kertz (1991)). *Suppose that venture capitalists are constrained by the time dynamics described above; moreover, suppose that the venture capitalists know the distribution F of deal values (which has non-negative support and a tail index $0 < \gamma < 1$) and follow an optimal selection strategy. Then (5) holds for the selection function* $\text{Select}(\cdot)$ *implied by the time dynamics, and moreover*

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[\text{Select}(\{X_1, ..., X_n\})\right]}{\mathbb{E}\left[\max\{X_1, ..., X_n\}\right]} = \frac{(1-\gamma)^{-\gamma}}{\Gamma(1-\gamma)},$$

where Γ is the gamma function.

This result also implies that the limiting cost of time dynamics only depends on the tail index γ. When $\gamma = 0.4$, then the described worst-case time dynamics reduce the expected return on investment of the venture capital firm by about 20%.

Appendix B Proof of Proposition 1

We proceed in two steps. First we show that the distribution of the V_i has the same tail index γ as F. Thus, an expression equivalent to (2) holds with the X_i replaced by V_i, and so the size of the best valuation scales at the desired rate:

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[\max\{V_1, ..., V_{[\alpha n]}\}\right]}{\mathbb{E}\left[\max\{V_1, ..., V_n\}\right]} = \alpha^\gamma \text{ for all } \alpha > 0. \tag{6}$$

Second, we need to show that as sample sizes grow, valuations remain somewhat representative of the value of the underlying deal. If, for i^* defined as in (4), we can show that $\mathbb{E}[X_{i^*}]/\mathbb{E}[V_{i^*}]$ converges to a finite limit as $n \to \infty$, then (5) follows directly from (6).

Now, under our conditions on the X_i and E_i, it can be shown that (e.g. Jessen and Mikosch 2006, Lemma 4.2)

$$\lim_{t\to\infty} \frac{\mathbb{P}[E \cdot X > t]}{\mathbb{E}\left[E^{1/\gamma}\right] \mathbb{P}[X > t]} = 1. \tag{7}$$

This formula directly implies that the V_i have tail index γ. Our second step requires some more work. We start with a well-known fact that can be verified using Potter's inequalities: If X has a tail index

$0 < \gamma < 1$, then

$$\lim_{t \to \infty} \frac{\mathbb{E}[X|X > t]}{t} = \frac{1}{1-\gamma}. \tag{8}$$

Now, using Bayes' rule, we can write (note that we can interchange limits and integration by the dominated convergence theorem because of Potter's inequalities and because G is bounded below by $B > 0$)

$$
\begin{aligned}
\lim_{t \to \infty} \frac{\mathbb{E}[X|E \cdot X > t]}{t} &= \int \lim_{t \to \infty} \frac{\mathbb{E}[X|X > t/\eta]}{t} \frac{\mathbb{P}[X > t/\eta]}{\mathbb{P}[E \cdot X > t]} \, dG(\eta) \\
&= \int \frac{1}{\eta(1-\gamma)} \lim_{t \to \infty} \frac{\eta^{1/\gamma} \mathbb{P}[X > t]}{\mathbb{E}[E^{1/\gamma}] \mathbb{P}[X > t]} \, dG(\eta) \\
&= \frac{1}{1-\gamma} \mathbb{E}\left[E^{\frac{1-\gamma}{\gamma}}\right] / \mathbb{E}\left[E^{\frac{1}{\gamma}}\right].
\end{aligned}
$$

By writing H for the distribution of $V = E \cdot X$ and $a(t) = \mathbb{E}[X|E \cdot X = t]$, we can re-write the above result as

$$\lim_{t \to \infty} \frac{1}{t(1-H(t))} \int_t^\infty a(t) \, dH(t) = \frac{1}{1-\gamma} \mathbb{E}\left[E^{\frac{1-\gamma}{\gamma}}\right] / \mathbb{E}\left[E^{\frac{1}{\gamma}}\right],$$

which, paired with (8), implies that

$$\lim_{t \to \infty} \frac{\int_t^\infty a(t) \, dH(t)}{\int_t^\infty t \, dH(t)} = \mathbb{E}\left[E^{\frac{1-\gamma}{\gamma}}\right] / \mathbb{E}\left[E^{\frac{1}{\gamma}}\right].$$

Finally, the maximum over m values of V_i has distribution H^m, and so we conclude that

$$
\begin{aligned}
\lim_{m \to \infty} \frac{\text{Select}\{X_1, ..., X_m\}}{\mathbb{E}[\max\{V_1, ..., V_m\}]} &= \lim_{m \to \infty} \frac{\int_0^\infty a(t) H^{m-1}(t) \, dH(t)}{\int_0^\infty t H^{m-1}(t) \, dH(t)} \\
&= \mathbb{E}\left[E^{\frac{1-\gamma}{\gamma}}\right] / \mathbb{E}\left[E^{\frac{1}{\gamma}}\right],
\end{aligned}
$$

leading us to the desired result.

Appendix C   The Growth Rate of the Extreme Order Statistics

In this section we verify that, as claimed in Section 4.3, the main results from Section 4.1 hold if the VC firm invests in the top k deals instead of only selecting the best deal. The main technical device

needed for generalizing our proofs is a result on the joint convergence of extreme order statistics. Let $X_{1,n}$, ..., $X_{k,n}$ be the top k out of n deals as described in Section 4.3. Now, using Theorem 2.1.1 and Proposition B.1.9.5 of de Haan and Ferreira (2006), we can check that for any fixed s an analogue to 2 holds: If the X_i are drawn from a distribution with tail positive support and a tail index $0 < \gamma < 1$, then $\lim_{n\to\infty} \mathbb{E}[X_{s,\alpha n}]/\mathbb{E}[X_{s,n}] = \alpha^\gamma$. This immediately implies that, for any fixed k,

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{s=1}^{k} X_{s,\alpha n}\right] \bigg/ \mathbb{E}\left[\sum_{s=1}^{k} X_{s,n}\right] = \alpha^\gamma.$$

The key ingredient here is Theorem 2.1.1; Proposition B.1.9.5 (or Potter's inequality) is only needed to go from convergence in distribution to convergence in moments. Using a similar argument, we can also establish analogues to the results presented in Appendix A in a setting where we select k deals instead of just 1.