The Science of Venture Bets:
Refining Portfolio Construction Through a Simulation-Optimization Framework

Mike Arpaia, Nic Lowden, Jonas Vetterle
Moonfire
{mike, nic, jonas}@moonfire.com

Abstract

As boundaries between Pre-Seed, Seed, and Series A continue to blur, traditional strategies for allocating venture capital across these early rounds often lack robust quantitative justification. In this paper, we propose a comprehensive framework for structuring a fund's portfolio construction across three critical sub-stages of modern seed investing: *Pre-Seed*, *Seed Inception*, and *Seed Expansion*. We employ a Monte Carlo simulator that models multi-stage lifecycles and power-law exit dynamics, then couple it with a Bayesian optimization routine to systematically identify high-performing allocation policies. Using rigorous simulation and adaptive experimentation, we provide a data-driven perspective on early-stage portfolio design that is particularly well-suited to today's larger, more competitive seed environment.

1 Executive Summary

This report applies a data-driven approach to portfolio construction at the earliest stages of venture capital, modeling allocations across *Pre-Seed*, *Seed Inception*, and *Seed Expansion* under a power-law exit environment. By combining Monte Carlo simulation and Bayesian optimization on millions of candidate strategies, we evaluate fund performance using simulated Distributions to Paid-In (DPI). Several key findings emerge:

1. **Distribute Capital Across Multiple Early Stages.** Diversifying initial bets among Pre-Seed, Seed Inception, and Seed Expansion broadens coverage and raises the likelihood of identifying "mega-winner" outliers.

2. **Limit Follow-On Allocations.** Allocating excessive capital to follow-ons erodes DPI by narrowing the overall funnel and blending the overall entry valuation. In our simulations, a minimal follow-on reserve per category yielded the strongest outcomes.

3. **Maintain a Moderate Discovery Co-Investment Bucket.** Allocating 5–20% to Discovery investments optimizes fund performance by expanding portfolio breadth in a capital-efficient way, increasing the likelihood of capturing an outlier. However, allocating capital well beyond this range diminishes returns as that expansion diverts capital from initial investments.

4. **Fund Size Matters Less Than Discipline.** Although larger fund sizes appeared optimal in several runs, moving within the $110M–$150M range had a smaller effect than adhering to low follow-ons and balanced stage allocations. Merely scaling the size of a fund up or down does not guarantee higher DPI if sourcing quality cannot be maintained. In general, our results indicate that *how* capital is deployed typically outweighs modest changes in total fund size.

Fund managers will see the best results by curbing follow-ons, diversifying early bets across the different categories of Pre-Seed and Seed, and selectively allocating capital to discovery investments. This balanced, stage-aware strategy is best suited to an environment where a small set of winners dominates outcomes.

Contents

2 Introduction

During the last decade, the landscape of early-stage venture capital (VC) has undergone dramatic changes in both the size and nature of seed funding. Today's "Seed" can range anywhere from a \$500k ticket for a first-time founder still validating an idea to a \$6m–\$10m round led by serial entrepreneurs with robust early traction, breaking the lines between what used to be clear distinctions between Pre-Seed, Seed, and Series A. As a result, founders face higher expectations and must often traverse an elongated timeline before achieving the metrics historically demanded at later funding rounds. Traditional heuristics for allocating capital in early-stage ventures, for example, "invest in 30 deals at the seed stage, save half for follow-ons", struggle to capture these new complexities.

We have observed that many founders now need multiple "seed stage" checks, some smaller, some substantially larger, before they reach the thresholds for a modern Series A round. In parallel, multistage investors and new entrants crowd the seed arena, inflating valuations and changing how quickly (or slowly) companies progress between milestones.

2.1 The Decomposition of "Seed"

The term "Seed" no longer represents a single, well-defined venture funding round. It can refer to an early \$500k ticket written before product launch (*Pre-Seed*), a \$2m–\$3m round backing a seasoned founder with a partially validated product (*Seed Inception*), or even a \$6m "Seed Expansion" round that is effectively bridging a startup's path to a \$10m Series A. This enormous range not only confuses founders and VCs, but also obscures the actual progress and risk profile of the company.

In our recent blog post *"Seed is broken. Here is how to fix it"*[1], we outlined a simplified scheme to describe how seed-stage companies truly differ. For many of the startups, we see:

- **Pre-Seed** (often under \$2m) supports a team that may be first-time founders or early builders just completing product discovery.
- **Seed Inception** (commonly \$3–\$5m, or more if the founders have a significant track record) offers enough capital for a real product build, first hires, and initial market traction.
- **Seed Expansion** (again \$3–\$6m or beyond) "tops up" successful pre-seed or inception rounds, bridging the gap toward the ever-larger Series A.

By distinguishing these sub-stages, we gain clearer insight into a startup's maturity and risk profile, and can allocate capital accordingly. In designing an early-stage venture fund portfolio construction, one of our core questions becomes: *How should we split our capital among these three sub-stages, balancing smaller, earlier bets with larger checks for high-conviction expansions?* Our thesis is that the extremely skewed (power-law) nature of VC outcomes implies that both wide funnel coverage (at Pre-Seed, Seed Inception, and Seed Expansion) and follow-on reserve can materially affect final returns.

2.2 Portfolio Construction Considerations

Complexity in Portfolio Design. Even with the refined definitions of Seed, portfolio managers still have to decide: *How many deals should we back in each sub-stage? How large should each check be? How much capital should we reserve for follow-on rounds?* In a world where the line from pre-product concept to mega-\$6m "seed" round has blurred, ignoring the stage-based differences can lead to suboptimal capital deployment. In addition, the small fraction of outliers that deliver outsized returns means that an investor's exact approach to stage selection, ownership targets, and follow-on investment can drastically change fund-level metrics.

[1] `https://www.moonfire.com/stories/seed-is-broken-heres-how-to-fix-it/`

Need for Quantitative Approaches. To tackle these complexities, *simulation* and *automated optimization* offer a much-needed complement to heuristic approaches. A carefully built simulation can capture the following.

- Different sub-stages of "Seed", from Pre-Seed to larger Seed Inception and Seed Expansion rounds.
- Varying probability of failure tied to the maturity of a startup.
- Potentially high valuations and increased valuation step-ups if the company gains traction.
- Significant dilution in follow-on rounds if the fund does not maintain pro rata ownership.
- The heavy-tailed nature of venture exits, where a small set of "home runs" skews overall returns.

An *optimizer* can systematically test allocation strategies. For example, how much of the fund to deploy at Pre-Seed versus Seed Inception, how much capital to allocate to follow-ons, etc. This paper demonstrates how such a simulation-optimization loop can yield deeper insights into stage allocation than unstructured guesswork or simplistic "fixed fraction" guidelines.

2.3 Organization of the Paper

The rest of the paper is structured to address both the contextual underpinnings of why a more refined breakdown of the seed stage is needed and the technical framework of how to build and optimize such a fund design:

- **Section 3** gives context to our approach by examining how early-stage deal sizes have evolved and why modern seed rounds increasingly require multiple funding segments. It also reviews empirical evidence for the power-law nature of VC exits, laying the groundwork for our simulation assumptions.
- **Section 4** provides a detailed rationale for modeling venture outcomes with a truncated power law. We discuss key properties such as tail exponents, probability mass concentration, and the practical upper bounds that justify a truncated approach. These considerations inform the random exit draws in our simulation.
- **Section 5** describes our core simulation methodology, including the multi-stage venture lifecycle, valuation step-ups, dilution mechanics, and exit timing. It outlines how capital calls are scheduled, how failures are accounted for, and how various scenarios (e.g., partial exits) can be incorporated if desired.
- **Section 6** introduces the Bayesian optimization strategy used to identify high-performing allocations. It explains how we formulate the optimization objective (DPI), define the parameter space (e.g., fund size, category splits, follow-on reserves), and iteratively refine allocations via an acquisition function.
- **Section 7** delves into the recurring themes observed across multiple simulation-optimization runs. We discuss how fund size, stage-specific splits, and follow-on budgets interact to drive outcomes in a power-law environment, while highlighting patterns of convergence and practical considerations for sourcing deals.
- **Section 8** synthesizes our principal findings regarding stage allocation, follow-on constraints, discovery investment strategies, and overall portfolio sensitivity. It also flags the limitations of our model and proposes future directions, such as dynamic rebalancing and sector-specific parameter tuning.
- **Appendix A** presents in-depth quantitative results from the Bayesian optimization trials. We analyze allocations that maximized DPI, compare them against alternative strategies, and illustrate how key parameters (e.g., follow-on fraction) influence performance under different assumptions.

- **Appendix B** details the software architecture for our simulation framework. It describes how the system is modularized into configuration objects, portfolio construction logic, and a runtime engine capable of orchestrating multi-stage deals, capital calls, and exit events at scale.
- **Appendix C** reviews how and why we use the Beta distribution to model holding periods in our simulation. We discuss how changing α and β parameters shifts the distribution of exit or follow-on timing from front-loaded to back-loaded, and how this flexibility allows us to match different observed patterns in real-world venture exits.

We now turn to Section 3 which aims to discuss the larger context on venture portfolio design and empirical evidence for the kinds of highly skewed outcomes that motivate our approach.

3 Background

Although the fundamental mechanics of early-stage venture investing have existed for decades, the past few years have seen a dramatic evolution in "Seed" round sizes, expectations, and investor competition. In this section, we first expand on the impetus behind distinguishing **Pre-Seed**, **Seed Inception**, and **Seed Expansion**—three sub-stages that we view as essential for understanding how capital can be effectively allocated in today's environment. We then revisit the empirical rationale for modeling exit outcomes with heavy-tailed distributions, an assumption that underlies our simulation approach.

3.1 Why Three Sub-Stages of Seed?

Pre-Seed. Historically, "Pre-Seed" was a small ticket, often under $2m, allocated to a newly formed or even pre-incorporation team. Founders of Pre-Seed startups might still be solidifying their product concept or performing initial market validation. Pre-Seed is thus the riskiest capital in the sense that the company has exhibited limited traction or external validation (if any). However, it is also a chance to secure meaningful ownership at a lower valuation. Funds that excel at sourcing and supporting talented founders early can capture enormous upside if the startup matures successfully.

Seed Inception. Between $3m–$5m, a "Seed Inception" round serves as the company's true "launch capital." The team typically has a prototype or an early MVP (Minimum Viable Product). These funds enable the hiring of key engineers and product talent, the refinement of the product-market fit, and the initial go-to-market strategy. For experienced entrepreneurs, or where the addressable market is very large, these checks can exceed $6m. In particular, the difference between a "Pre-Seed Startup" and a "Seed Inception Startup" could be measured just in 9–12 months of extra product traction.

Seed Expansion. Often, even a $3–$4m initial seed is not enough runway in today's market to reach Series A metrics. Consequently, many teams raise a second or "expansion" seed, topping up another $3–$6m to continue iterating on go-to-market strategy, to solidify revenue streams, and to scale the team. In addition, especially experienced founders can often make a compelling case for raising more growth capital earlier in the company's lifecycle. For an investor, *Seed Expansion* can often be an attractive investment opportunity when the founders have validated their product with real customers or are demonstrating especially promising early traction.

3.2 Implications for Fund Design

By explicitly recognizing these three sub-stages, we *decompose* what used to be a single "Seed" allocation into separate capital buckets. Each bucket has different check sizes, success probabilities, ownership targets, and follow-on dynamics. **A key hypothesis we tested was whether the relative proportions assigned to Pre-Seed vs. Seed Inception vs. Seed Expansion would non-trivially influence overall fund performance.**

- Too little Pre-Seed means missing out on opportunities that are extremely promising at their lowest valuation.
- Too little Seed Inception might under-allocate to highly investable early ventures that are just starting to show market validation.
- Too little Seed Expansion capacity could leave the fund unable to invest in seed rounds where the founders are exceptionally experienced and/or when the company is exhibiting especially promising early traction.

In addition, these strategic splits must accommodate the heavy-tailed nature of the outcomes, in which a single outlier can return the fund.

4 Mathematical Foundations of the Truncated Power Law

Venture capital is often characterized by a small fraction of deals generating disproportionately high returns, while the majority produce modest or zero returns. This section first examines the occurrence of such skewed outcomes, then formalizes how a *truncated power law* can be used to model exit multiples. We conclude by explaining why we keep these distribution parameters fixed when comparing different portfolio strategies.

4.1 Rationale for Power-Law Distributions

A recurring theme in venture portfolios is extreme *outcome concentration*. In typical venture portfolios, only a few positions account for the majority of the total gains. Factors like network effects, first-mover advantages, or winner-take-all market structures often push certain companies to achieve outsized growth, skewing the overall returns.

It is generally well understood that these large exits are more common than a normal or even a lognormal model would predict. In other words, the "tail" of the distribution is *heavier*. Power-law or Pareto-type fits,

$$P(X > x) \ \propto \ x^{-\alpha},$$

better capture the observed frequency of very large outcomes such as $50\times$ or $100\times$ returns. Although these super-exits remain rare, they occur more often than would be expected under thinner-tailed models.

However, in reality, no startup's value can grow without limit. Even extraordinarily successful IPOs or acquisitions have practical upper bounds, whether determined by addressable market sizes, competitive dynamics, or other factors. Consequently, an unbounded power law is not fully realistic, prompting the use of a *truncated* version. This truncated approach still acknowledges the possibility of large multiples, but tapers off probabilities at some maximum feasible exit level.[2]

A key parameter in these models is the *exponent* α. A smaller α (say 1.5–1.8) implies heavier tails, meaning ultra-large outcomes remain relatively more probable. A bigger α (2.5–3.0) makes home-run exits rarer and imposes a steeper probability drop-off beyond the typical valuation range. From a practical standpoint, knowing the tail exponent is critical to understanding how often "unicorn-plus" returns might appear.

Such heavy tails also interact strongly with multi-stage investment strategies. Because only a few deals may drive most of the gains, the decision of which winners to back in follow-on rounds can make or break fund performance. In our simulation, we incorporate the power-law exit model into a multi-stage framework to see how the concentration of capital in potential outliers—and the ability to maintain ownership—affects overall returns.

One might speculate that more skilled or sector-focused investors could alter the shape of the power-law distribution itself, either by lowering α (so that mega-hits are more common) or by raising the maximum possible exit. Although this is plausible, the precise link between "competence" and distribution shape is far from well understood, and there is no compelling evidence to support the idea that skill simply shifts the tail. Consequently, for the purpose of our simulation environment, we keep the distribution parameters *fixed* during simulations to avoid confounding the effect of investor decision-making (capital allocation, follow-on strategy, etc.) with changes in the underlying exit environment.

4.2 Mathematical Formulation and Model Parameters

To formalize these ideas, let X be the exit multiple of invested capital for a given company (so $X > 1$ indicates a profitable exit, and $X < 1$ implies a loss). If exits were distributed by an unbounded power law with exponent $\alpha > 1$, the probability density function (PDF) would be:

$$f(x) \ \propto \ x^{-\alpha}, \quad x \geq x_{\min}.$$

[2] For more context on the truncated power law and it's application to venture capital portfolio simulation, please also refer to the report that we published in 2023 which explored this topic in a more narrow set of experiments: `https://arxiv.org/pdf/2303.11013`.

In practice, valuations do not extend infinitely, so we define a *truncated* power law on the interval $[x_{\min}, x_{\max}]$. The PDF then becomes

$$f(x) \;=\; \frac{(\alpha - 1)\, x^{-\alpha}}{x_{\min}^{1-\alpha} - x_{\max}^{1-\alpha}}, \quad x_{\min} \le x \le x_{\max},$$

with $f(x) = 0$ outside that interval. The normalization factor in the denominator ensures the PDF integrates to 1.

Typical power-law parameters in venture capital contexts range roughly from $\alpha \approx 1.5$ to 3.0. The minimum value $x_{\min}$ might be close to 0.1 or 0.3 to accommodate near-total write-offs, while the maximum $x_{\max}$ could be set anywhere from 100 to 1,000 to represent potential "mega-exits." In our own simulator, we use default values of

$$\alpha = 2.05, \quad x_{\min} = 0.35, \quad x_{\max} = 600.$$

This choice reflects a moderately heavy tail, enough to allow occasional extreme outcomes yet still keep returns within a plausible upper range. By fixing these parameters for all simulation runs, we avoid mixing changes in the *return environment* (i.e. the power-law shape) with changes in *investment strategy*. If desired, future versions of the model could relax this assumption to examine scenarios where skillful investors systematically enjoy heavier tails, but that would introduce a separate dimension of variability outside the scope of our current comparative analysis.

5 Simulation Methodology

This section provides a detailed account of our venture capital (VC) simulation methodology, from the initial capitalization of a fund to the terminal outcomes of individual portfolio companies. Our overarching objective is to generate a representative *Monte Carlo* environment that captures the uncertainties, high skewness, and multi-stage nature of early-stage venture investing. Although the simulation is extensible to a variety of fund sizes, strategies, and markets, the discussion here focuses on the core mechanics: how investments are created, how they progress through funding rounds, how valuations evolve, and how exits ultimately impact portfolio-level metrics.

5.1 Overall Framework and Key Inputs

At a high level, the simulation revolves around three main components:

1. **Fund Setup:** We initialize a virtual VC fund with a specified size, management fees, and an investment period over which new deals (initial checks) can occur. Any portion of the fund that exceeds management fees is treated as *investable capital* and subdivided across multiple *categories* (e.g., pre-seed, seed, follow-ons).

2. **Deal Generation and Progression:** The simulation generates "initial deals" for each category within the investment period, spacing them out in pseudo-random fashion. Each deal references an *initial* check (ticket size and ownership) and, potentially, follow-on checks. Companies raised through these deals evolve or fail across a user-defined set of stages (Pre-Seed, Seed, Series A, . . .). At each stage transition, we apply a probability of failure, a valuation uplift for successful progression, and track the possibility of investor follow-ons.

3. **Exit Events and Performance Measurement:** Companies that survive the final stage (or at any stage, if so configured) exit according to a distribution of holding periods. The ultimate returns are modeled as random draws from a truncated power-law, reflecting the highly skewed nature of startup outcomes. The simulation computes aggregated fund metrics such as IRR, DPI, TVPI, and MOIC, repeating the entire process multiple times to obtain robust statistical estimates.

By default, the simulation loops over all companies in the portfolio once the final investment horizon is reached, applying the exit logic for each surviving position. However, users can modify this approach to accommodate partial exits at intermediate stages or alternative exit-timing heuristics.

5.2 Initial Fund Capital and Category Allocations

Before any deals are created, we define the *gross fund size*, including any management fees. The actual capital available for investment is typically computed as

$$\text{Investable Capital} = \text{Fund Size} \times (1 - \text{Management Fees}).$$

For example, a \$120 million fund with 23.25% in fees (over the course of the fund's lifetime) might have approximately \$92 million to invest in actual transactions. This investable amount is then subdivided among a set of categories. Each category might correspond to:

- Early vs. Late Rounds (Pre-Seed, Seed Inception, Seed Expansion, etc.)
- Special Discovery Investment Buckets

Depending on the strategy, one may assign distinct ticket-size ranges or follow-on strategies to each category. For example, a Pre-Seed category might limit initial checks to \$0.8–\$1.2 million, aiming for a 10–15% ownership stake, whereas a Seed category might use \$2–\$3 million initial checks. The user can configure these budget fractions and ticket-size distributions in advance. The simulation enforces the constraint that the sum of all category allocations equals approximately[3] 100% of the invested capital.

5.3 Deal Flow Generation and Capital Calls

One of the core tasks of the simulation is to *schedule* when initial deals take place within the investment period of the fund (e.g., 2.5 to 3 years). Rather than place deals uniformly, we typically employ an exponential or Poisson-like inter-arrival mechanism. Concretely:

1. Let N be the total number of initial deals desired across all categories (derived from each category's budget and average ticket size).
2. Convert the investment horizon T (in years) to days, obtaining $365\,T$.
3. Draw N exponentially distributed inter-arrival times so that on average we reach N deals over the $365\,T$ days.

By sorting the cumulative sums of these inter-arrival times, the simulation obtains a realistic schedule of deals. Each category obtains its fair share of deals based on how many it is allocated. For example, if "Pre-Seed" is allotted 20 initial deals while "Seed Inception" is allotted 40, the model uses random sampling within each category to generate those 20 or 40 deals.

Quarterly Capital Calls. To reflect the notion that funds typically call capital in batches, the simulation groups deals into quarterly windows. Each quarter, the total ticket sizes of all deals scheduled within that window are summed, and if enough *un-called* capital remains, it is *called* from the limited partners. If insufficient capital remains, some deals may be skipped entirely, capturing the real-world constraint of capital exhaustion. This process ensures that not all capital is available on day one and fosters a more granular cash-flow timeline. These more realistic cash flows can be used to calculate performance metrics such as IRR.

5.4 Stage Transitions: Failure and Valuation

After an initial investment is recorded, each company attempts to "move" through a predefined sequence of stages (e.g., Pre-Seed → Seed → Series A → ...). Two essential dynamics govern these transitions: *failure probability* and *valuation uplift*.

5.4.1 Failure Probability and Signalling Risk

For each stage transition (e.g., from Pre-Seed to Seed), we assign a probability that the company will fail to advance. For example, one might assume the following:

Stage Transition	Failure Probability
Pre-Seed → Seed	0.20 (20%)
Seed → Series A	0.15 (15%)
Series A → Series B	0.12 (12%)
Series B → Series C	0.08 (8%)
Series C → Series D	0.05 (5%)

Table 1: Stage Transitions and Associated Failure Probabilities

If a given company attempts to raise a new round and fails, its value is written down to zero (immediate exit at $0). By default, the model interprets "failure" as an inability to secure follow-on financing or a fundamental business collapse. This event ends the simulation for that company.

[3]Given the nature of large investment round sizes, it can be difficult to construct a portfolio construction which allocates exactly 100% of the investable capital. As such, we allow a minimal amount of flexibility for the optimizer.

Furthermore, we incorporate *signalling risk*: if a fund previously held a large stake (e.g., above 8% ownership) but declines to follow on, the market might perceive this as negative information, thus increasing that transition's failure probability. Thus, if the base failure rate is p_{fail}, signaling risk might multiply it by some factor (e.g., 1.25). This approach captures the notion that other investors may be disincentivized to invest if a lead investor refuses to reinvest.

5.4.2 Valuation Uplift and Ownership Dilution

If a company survives its previous stage, its valuation increases by a specific multiplier at each successful round:

Stage Transition	Valuation Uplift
Pre-Seed → Seed	2.0×
Seed → Series A	3.0×
Series A → Series B	2.5×
Series B → Series C	2.0×
Series C → Series D	1.8×

Table 2: Stage Transitions and Associated Valuation Uplifts

This factor is meant to approximate how valuations typically escalate with each successful round. After applying the multiplier, the model infers how much *new capital* is raised. If existing investors do not match their pro rata ownership in that new capital, they experience *dilution*, effectively shrinking their ownership fraction by some median proportion (e.g., 20% for a typical Seed → Series A round).

It's important to note that these stage transition multiples are used for follow-on and dilution modeling but not for determining final exit values. Instead, exit outcomes are drawn from a truncated power-law distribution (see Section 5.5). This approach more effectively captures the skewed nature of venture returns, where a few outliers drive most of the value. While stage multiples help approximate ownership dynamics between rounds, they assume a more linear growth path, whereas the power law better reflects the extreme variance in actual exit outcomes.

The combination of failure, valuations, and dilution significantly influences the eventual realized value of each position, especially if the fund invests pro rata only in certain "winners". Over multiple stage transitions, a single strong performer might balloon in value, offsetting numerous write-offs—mirroring the well-known "power law" dynamic of VC portfolios.

5.5 Exit Timing and Distributions

Eventually, each surviving company undergoes an exit event. In our baseline approach, we assume the following:

1. **Holding Period:** The simulation samples a random holding period H (in years) from a Beta distribution on $[t_{\min}, t_{\max}]$. For example, if $t_{\min} = 4$ and $t_{\max} = 10$, the code transforms a Beta(α, β) draw $X \in [0, 1]$ into

$$H = t_{\min} + X\,(t_{\max} - t_{\min}).$$

   Larger values of α in the Beta distribution skew exits to occur later in that range, while smaller α might cluster them earlier.

2. **Exit Multiple:** Once the company is scheduled to exit, a *truncated power-law* random variable yields the final multiple of the invested capital. The probability density function (PDF) on $[x_{\min}, x_{\max}]$ is given by

$$f(x) = \frac{(\alpha - 1)\,x^{-\alpha}}{x_{\min}^{1-\alpha} - x_{\max}^{1-\alpha}}, \quad x_{\min} \le x \le x_{\max},$$

with $\alpha > 1$. A smaller α indicates a heavier tail (more chance for extreme outliers). The product of "total invested" and the sampled multiple becomes the final distribution to the fund.

3. **Distribution of Proceeds:** We treat each exit as a full liquidation. If partial or staged exits are desired, the user can incorporate advanced logic to schedule partial liquidity events or secondary share sales at intermediate times.

At the portfolio level, these randomly timed and sized exits drive the computation of standard VC performance metrics. Because multiple companies exit at different times, the final realized distribution is the sum of many stochastic outcomes, with a few "unicorn" or "mega" multiples potentially dominating the final results.

5.6 Computing Portfolio Metrics

The simulation computes four principal metrics after all companies have either failed or exited:

- **IRR (Internal Rate of Return):** Incorporates the exact timing of capital calls (negative cash flows) and exits (positive cash flows). If D_t is the net inflow or outflow on date t, the IRR r solves
$$\sum_{t=1}^{T} \frac{D_t}{(1+r)^{(t-t_1)/365}} \;=\; 0,$$
where we measure t in days (or any consistent unit). An IRR of 0% means that the fund returned exactly its invested capital (in NPV terms), while a higher IRR indicates a more rapid capital recovery.

- **DPI (Distributions to Paid-In):** Ratio of total capital returned (realized distributions) over total paid-in capital. This metric ignores unrealized value, focusing on *actual* cash back to LPs.

- **TVPI (Total Value to Paid-In):** Considers both realized distributions and any remaining *unrealized* value in the portfolio, divided by the total paid-in capital. Thus,
$$\text{TVPI} \;=\; \frac{\text{Realized} + \text{Unrealized}}{\text{Paid-In}}.$$
A TVPI above 1.0 implies potential gains, though they are not guaranteed until those unrealized positions exit.

- **MOIC (Multiple on Invested Capital):** Measures how much value has been generated by the invested capital, often represented as a simple multiple (1.8×, 2.5×, etc.).
$$\text{MOIC} = \frac{\text{Realized} + \text{Unrealized}}{\text{Invested Capital}}$$

By running the entire simulation multiple times, each with different random seeds, one obtains a distribution of outcomes for IRR, DPI, etc. Analysts can then measure the *mean*, *median*, or *value-at-risk (VaR)* of these metrics, or compare multiple proposed strategies within the same environment.

5.7 Calibration, Sensitivity, and Extensions

A key advantage of this approach is its *flexibility* in calibrating the underlying assumptions:

- **Failure Probabilities:** Adjusting p_{fail} at each stage can reflect harsher or more generous funding environments, or differences across geographies (e.g., US vs. Europe).

- **Valuation Uplifts:** Changing u_i between rounds modifies the typical step-up in post-money valuations. Lower uplifts yield more conservative valuations, while higher ones reflect "frothier" markets.
- **Dilution Percentages:** The fraction of new equity sold at each round can be increased to mirror larger rounds or decreased for smaller, more incremental raises.
- **Exit Timing (Beta Distribution):** Altering α, β or the range $[t_{\min}, t_{\max}]$ can shift the typical hold period for investments, capturing either quick M&A cycles or slower, more R&D-intensive industries.
- **Exit Multiples (Truncated Power Law):** The exponent α, lower bound $x_{\min}$, and upper bound $x_{\max}$ can be tuned to represent different risk appetites or historical track records. A heavier tail (α closer to 1) raises the odds of a "mega" outcome.

In principle, one could also incorporate partial exits, bridging rounds, or advanced signaling risk logic. The main idea is that this simulation environment is *modular* enough to handle these complexities while preserving the same fundamental flow.

6 Optimization Methodology and Rationale

This section presents an account of how we optimize the design of a venture capital (VC) fund's allocation strategy. Our main objective is to find a parameter vector that maximizes a given performance metric, here the *Distributions to Paid-In* ratio (DPI). We begin by outlining how these parameters relate to our underlying portfolio construction (e.g., pre-seed vs. seed allocations), and then describe in detail how we embed a Bayesian optimization loop on top of repeated Monte Carlo simulations.

6.1 Investment Categories

In our portfolio construction model (Sections 5 and B), we define investment "categories" corresponding to different stages or types of deals: *Pre-Seed*, *Seed Inception*, *Seed Expansion*, and *Discovery.*

Each category has a fraction of the total *investable capital* (i.e., net fund size after fees) assigned to it, and these fractions determine how many deals and how large the checks can be. In addition, each fraction is split further between *initial* investments and *follow-on* rounds. Hence, the fundamental question is: *How do we choose these category fractions and the initial/follow-on split to maximize long-run returns?*

6.2 Parameter Space and Fund-Level Variables

To encode the above decisions systematically, we introduce a parameter vector

$$\Theta \;=\; \bigl(F,\, \theta_{\text{pre}},\, \theta_{\text{inc}},\, \theta_{\text{exp}},\, \theta_{\text{co}},\, \rho\bigr).$$

Below we define each component in more depth:

1. **Fund Size** (F): A discrete choice (e.g., {110M, 115M, 120M, 125M, 135M, 140M, 145M, 150M}), which is the total committed capital of the fund. Our simulation subtracts management fees to yield the *investable capital*, but F itself can vary so that we can test performance across different scale scenarios.

2. θ_{pre}, θ_{inc}, θ_{exp}: Each lies in $[0, 1]$ and represents a *raw* weighting for the three main categories:

   - θ_{pre} for Pre-Seed,
   - θ_{inc} for Seed (or Seed Inception),
   - θ_{exp} for Seed Expansion.

   However, these θ-values do not directly sum to 1.0; we apply a softmax transformation (detailed below) to turn them into actual allocation *fractions* for each category.

3. θ_{co}: Also in $[0, 1]$, controlling how much capital is allocated to the **Discovery** category. Conceptually, this discovery investment fraction is determined relative to the main categories' fractions, since we typically want discovery investments to be smaller than the main categories or bounded in some proportion to them (e.g., we do not want discovery investments to overshadow the main allocations).

4. $\rho \in [0, 1]$: The fraction of each category's budget earmarked for *initial* vs. follow-on checks. We usually map ρ into a narrower interval such as [0.5, 0.95] to ensure that at least half the category's budget is used for initial deals and at most 95%.

Once we have the final category fractions, each category's share of the fund is *further subdivided* into initial and follow-on budgets via ρ. Together, these parameters (F, $\theta_{\dots}$, ρ) define a coherent plan for how capital is distributed across the entire life cycle of the fund.

6.3 Transformations: From Raw Parameters to Category Fractions

Softmax for Main Categories. The raw triple $(\theta_{\text{pre}}, \theta_{\text{inc}}, \theta_{\text{exp}})$ is turned into normalized weights via

$$\left(\alpha_{\text{pre}}, \alpha_{\text{inc}}, \alpha_{\text{exp}}\right) \;=\; \text{softmax}\Big(\theta_{\text{pre}}, \theta_{\text{inc}}, \theta_{\text{exp}}\Big),$$

where

$$\alpha_{\text{pre}} \;=\; \frac{\exp(\theta_{\text{pre}})}{\exp(\theta_{\text{pre}}) + \exp(\theta_{\text{inc}}) + \exp(\theta_{\text{exp}})},$$

and similarly for α_{inc} and α_{exp}. We now have three fractions that sum to 1.0, describing how to distribute *all* capital among the main categories—*before* discovery investment is considered.

Minimum Category Fractions. We may require that each main category has at least a minimum fraction, say 5%, to ensure diversification. If any α_c from the softmax is below 5%, we may set $\alpha_c = 0.05$ and renormalize. This step prevents the optimizer from collapsing a category to a near-zero fraction (which might be locally optimal but unrealistic for a broad investment mandate).

Discovery Fraction. Next, θ_{co} in $[0, 1]$ controls how *large* discovery investment is relative to these main categories. One approach is to define

$$\alpha_{\text{co}} = \max\Big(0.05,\; \big(\theta_{\text{co}} \times 0.9\big)\; \min\{\alpha_{\text{pre}}, \alpha_{\text{inc}}, \alpha_{\text{exp}}\}\Big)$$

so that discovery investment is anchored to the smallest main category fraction but is never below some floor (e.g., 5%). After computing α_{co}, we reduce $(\alpha_{\text{pre}}, \alpha_{\text{inc}}, \alpha_{\text{exp}})$ proportionally so that the *sum* remains 1.0. This ensures that all four fractions, plus discovery investment, add up to 100% of the investable capital.

Initial vs. Follow-On Splits. Finally, each category fraction α_c is split via ρ. If ρ is mapped to $[0.5,\, 0.95]$, then for each category c:

$$\alpha_c^{(\text{init})} = \alpha_c\, \rho, \quad \alpha_c^{(\text{follow})} = \alpha_c\, (1 - \rho).$$

Hence, if $\rho = 0.8$, then 80% of category c's capital is designated for *initial* checks, and 20% for follow-ons. Altogether, we end up with 7 effective fractions (3 main categories $\times 2$ ways to split, plus discovery investment initial) that determine how the fund invests capital in different stages and at different points in a company's life.

6.4 Objective Function: Expected DPI via Monte Carlo

Distributions to Paid-In (DPI). Recall that DPI is defined as

$$\text{DPI} = \frac{\text{Total Capital Distributed (i.e., returned to LPs)}}{\text{Total Capital Paid-In}}.$$

A larger DPI means we have recovered more money relative to what we spent (neglecting unrealized value). Because the simulation includes random company outcomes, random exit multiples, and random follow-on participation, DPI for a given set of parameters Θ is a *random variable.*

Estimating Expected DPI. To approximate $\mathbb{E}[\text{DPI} \mid \Theta]$, we run the entire portfolio simulation n_{runs} times with different seeds (thus different random draws) but the same Θ. Each run produces a realization $\text{DPI}_j(\Theta)$ for $j = 1, \ldots, n_{\text{runs}}$. The average,

$$\widehat{\text{DPI}}(\Theta) = \frac{1}{n_{\text{runs}}} \sum_{j=1}^{n_{\text{runs}}} \text{DPI}_j(\Theta),$$

serves as our estimator. By making n_{runs} large enough (often 1,000 or 10,000), we reduce the Monte Carlo error, increasing confidence in which Θ yields superior DPI.

Maximizing $\widehat{\mathrm{DPI}}$. Our optimization target is thus:

$$\Theta^* = \underset{\Theta \in \mathcal{X}}{\operatorname{argmax}}\, \widehat{\mathrm{DPI}}(\Theta),$$

where $\mathcal{X}$ enforces F from the discrete set, the raw $\theta_{\dots} \in [0,1]$, $\rho \in [0,1]$, and all our min-fraction constraints. We solve this problem via Bayesian optimization as outlined below.

6.5 Bayesian Optimization Process

Motivation. Without additional structure, searching $\Theta \in \mathcal{X}$ might be done by random sampling or a naive grid. But we risk running thousands of simulations without systematically honing in on a better design. Bayesian optimization provides a structured approach: it builds a *surrogate model* of $\widehat{\mathrm{DPI}}(\Theta)$, uses an *acquisition function* to propose new Θ values, and iteratively refines its model.

Explicitly, Bayesian optimization allows us to run more efficient and targeted simulations, focusing on the most relevant parts of the parameter space and avoiding wasted effort. This means the modeling is more efficient, allowing us to reach an accurate conclusion more quickly, and ultimately leading to a better-optimized portfolio construction.

Surrogate Function. After each evaluation of $\Theta^{(t)}$ (i.e., after computing $\widehat{\mathrm{DPI}}(\Theta^{(t)})$ from n_{runs} simulations), we update a probabilistic approximation $m^{(t)}$ that predicts DPI for any $\Theta \in \mathcal{X}$. Commonly, $m^{(t)}$ might be a Gaussian Process (GP) or a tree-based regressor. The key is that $m^{(t)}$ tracks both the *expected* DPI at each point and the *uncertainty* in that estimate.

Acquisition Function. We then propose

$$\Theta^{(t+1)} = \underset{\Theta \in \mathcal{X}}{\operatorname{argmax}}\, a\big(\Theta \mid m^{(t)}\big),$$

where $a(\cdot)$ might be *Expected Improvement*, *Upper Confidence Bound*, or another strategy that balances *exploitation* (testing near known good solutions) and *exploration* (looking in uncertain regions). We evaluate $\widehat{\mathrm{DPI}}(\Theta^{(t+1)})$ via new simulations, feed the result back into $m^{(t+1)}$, and repeat.

Stopping. Since each step involves many Monte Carlo runs, we place a cap `max_trials` (e.g., 10,000). We can also stop if no improvement in best-found DPI occurs for, say, 250 trials in a row.

6.6 Putting It All Together

Algorithmically:

1. **Initialize**:
   - Set up $\mathcal{X}$: discrete F choices, $\theta_{\dots} \in [0,1]$, min fraction constraints, etc.
   - Launch a parallel worker pool for repeated simulation calls.
   - Create the Bayesian optimization engine with Objective = maximize DPI.
   - Set bestDPI $\leftarrow -\infty$.
2. **Loop over** $t = 1, \dots,$ `max_trials`:
   (a) *Acquire next point*:
   $$\Theta^{(t)} \;=\; \operatorname{argmax}_{\Theta \in \mathcal{X}}\, a\big(\Theta \mid m^{(t-1)}\big).$$
   (b) *Map raw* $\theta_{\text{pre}}, \theta_{\text{inc}}, \dots$ to final fractions $\alpha_{\text{pre}}, \dots, \alpha_{\text{co}}$, and split by ρ for initial vs. follow-ons.
   (c) *Compute* $\widehat{\mathrm{DPI}}(\Theta^{(t)})$:

- For $j = 1, \ldots, n_{\text{runs}}$, simulate the entire fund: generate deals, invest capital across categories, sample failures and exits, compute $\text{DPI}_j(\Theta^{(t)})$.
- Take average: $\widehat{\text{DPI}}(\Theta^{(t)}) = \frac{1}{n_{\text{runs}}} \sum_j \text{DPI}_j(\Theta^{(t)})$.

(d) *Update surrogate* $m^{(t)}$ with $\big(\Theta^{(t)}, \widehat{\text{DPI}}(\Theta^{(t)})\big)$.

(e) *Check improvement*: if $\widehat{\text{DPI}}(\Theta^{(t)}) > \text{bestDPI}$, record it and reset any "no improvement" counters.

(f) *Stopping checks*: if we exceed the maximum number of trials or if no improvement occurred for more than the maximum number of allowed steps, stop.

By iterating this process, we adaptively refine a data-driven model of how each parameter affects the expected DPI, gradually zeroing in on a better (and eventually best) design for the fund.

7 Results & Observations

In an effort to understand the implications of the results of our optimization process, we ran the full simulation optimization process several times. The results of a single representative run of the optimization process are described in detail in Appendix A.

That being said, the process of running the simulation-optimization multiple times revealed several recurring patterns and we believe that it is more important to focus on these higher-level patterns rather than just the outputs of a single optimization process:

1. **Low Follow-On Allocations Are Systematically Favored.** Across nearly all simulation runs which performed well, the fraction of capital reserved for follow-ons remained on the lower end of the tested range. Reserving too much for subsequent investments reduces initial coverage of new deals and, under power-law assumptions, depresses the probability of hitting a "mega-winner." In addition, larger follow-ons often blend up the fund's overall entry valuation, eroding potential multiples. Keeping follow-ons modest (often well below 20% per category) consistently led to stronger DPI outcomes.

2. **Maintain Consistency in Category Spreads.** Despite small fluctuations in the exact percentages for *Pre-Seed*, *Seed Inception*, *Seed Expansion*, and *Discovery*, the simulations routinely converged on a relatively balanced distribution of capital across these categories. In other words, devoting the bulk of investable capital to just one or two ticket types rarely emerged as optimal, indicating that multiple entry points help capture the rare outliers that dominate fund returns.

3. **Fund Size Exhibits Minimal Impact.** While we examined multiple fund sizes (e.g. \$110M–\$150M), variations in Distributions to Paid-In (DPI) across this range proved modest. Once a fund surpasses a threshold necessary to diversify effectively and maintain a few follow-on checks, simply adding or removing \$10M–\$20M of total committed capital did not materially alter expected performance.

Taken together, these findings underscore that while the simulation can fine-tune specific allocation percentages, the *broad lessons remain consistent*:

- Avoid over-reserving for follow-ons.
- Maintain a relatively even spread across relevant ticket categories.
- Do not overemphasize marginal differences in fund size.

This suggests that portfolio construction, in early-stage venture, may be *less sensitive* than is often assumed—*provided* the fund invests widely, stays stage-focused, and is disciplined about concentrated follow-ons.

7.1 Follow-On Capital Allocation Considerations

Although follow-on rounds can, in principle, help retain or increase ownership in promising companies, our simulations consistently indicate that large follow-on budgets lead to lower average DPI. We highlight two key drivers:

- **Breadth vs. Over-Concentration.** A broad funnel of initial investments is essential in a power-law environment, where a small number of outlier successes dominate aggregate returns. Reserving too much capital for later checks diminishes the number of new deals the fund can make, reducing the chance of capturing a mega-winner. Thus, large follow-on allocations often depress overall DPI simply by limiting the initial deal flow.
- **Entry Valuation Dilution.** Follow-ons typically occur at higher valuations, which reduces the multiple that can be realized upon exit. The more capital earmarked for these pricier

rounds, the more diluted the total return becomes—particularly in scenarios where the subsequent step-ups fail to materialize as expected. Despite the perceived information advantage of a follow-on investment bestowing a higher probability of investment return, it does not seem like that increased probability of return outweighs the higher cost of the investment (in terms of the entry valuation). Our simulation results show that raising the follow-on fraction can quickly erode final DPI.

In practice, a moderate approach appears most effective: reserve enough follow-on capital to double down on the most compelling opportunities, yet avoid locking up so much that it hinders broad initial coverage. Our findings support keeping follow-on reserves in a restrained range in order to balance incremental ownership upside with the need to cast a wide net at the outset.

7.2 Discovery Investment Observations

In our simulations, a modest discovery investment allocation (typically 5–20% of the fund) frequently emerged in top-performing portfolios. While this was not always a strict requirement for high DPI, it served as a flexible "reserve" that allowed opportunistic participation in deals that did not precisely align with the main ticket categories (Pre-Seed, Seed Inception, or Seed Expansion).

At higher proportions of discovery investment (above 20%), performance often dipped, probably because a too large bucket detracted from the purposeful coverage of the early stage. Conversely, too small or nonexistent a discovery investment allocation reduced the fund's agility to capitalize on outside-the-thesis or out-of-cycle deals that showed strong potential upside.

In practice, maintaining a discovery investment fraction near 5–20% appears beneficial for balancing the fund's primary mandate with the capacity to invest in exceptionally high-conviction opportunities.

7.3 Trial-Level Variability and Convergence

In many of our runs, the Bayesian optimization converged toward a similar region of parameter space, but the speed and path of convergence showed considerable variability. Early trials typically spanned a wide range of configurations: some favored extremely large follow-on budgets (over 30%), while others attempted concentrated deployments in just one or two ticket categories. Although these extremes rarely yielded strong DPI outcomes, they were valuable for probing boundaries of the parameter space. After 50–100 trials, the optimizer began to cluster around more balanced strategies, especially with regard to how initial checks were distributed among Pre-Seed, Seed Inception, and Seed Expansion. The shape of this convergence remained consistent across repeated runs of the process, albeit with slight differences in how many trials were required before the best allocations emerged.

An important byproduct of these repeated optimizations was evidence that localized maxima exist within the broader parameter space. In other words, two different allocations (e.g., 18% vs. 20% in Pre-Seed) can yield comparable DPI values over thousands of Monte Carlo simulations. Such close results highlight that exact percentages are less critical than adhering to the broad themes identified. Critically, allocating across stages, preserving moderate follow-ons, and ensuring enough initial coverage to sample potential outliers effectively.

7.4 Summary of Observations

In aggregate, these observations reinforce the central theme that balanced allocations across Pre-Seed, Seed Inception, and Seed Expansion, coupled with modest follow-ons, tends to produce robust and repeatable returns in a power-law exit environment. The convergence patterns across multiple optimization runs, the comparison with single-stage approaches, and the analyses of risk-return trade-offs all point to the same conclusion. While the exact fractions can vary slightly in high-performing portfolios, the broader guidelines of moderate discovery investments, minimal follow-ons, and a wide spread of initial checks consistently drive superior DPI outcomes.

Balancing all these considerations in real-world venture environments will inevitably be more nuanced, given sourcing constraints and unpredictable shifts in market sentiment (among other difficulties). Nevertheless, these results provide a clearer sense of the trade-offs and dynamics behind each capital allocation decision, from the distribution of deal types to the pacing of follow-ons. The next step for practitioners is to tailor these guidelines to their specific sector focus, network advantages, and risk appetite, while retaining the simulation-based insights on how crucial it is to keep coverage wide and follow-on commitments limited.

8 Conclusion

This paper introduced a simulation-optimization framework to explore multi-stage early-stage VC allocation policies under power-law exit assumptions. Our primary objective was to identify capital splits across *Pre-Seed*, *Seed Inception*, *Seed Expansion*, and *Discovery* categories that maximize the fund's Distributions to Paid-In (DPI). Repeated Monte Carlo simulation coupled with Bayesian optimization led to a few key conclusions:

- **Minimize Follow-On Allocations.** Small reserve budgets consistently outperformed more aggressive follow-on strategies. Because outliers govern fund returns, broad initial coverage raises the probability of capturing a mega-winner. Excessive follow-ons dilute this coverage and diminish overall multiples by raising the average entry valuation.
- **Balance Distribution Across Ticket Types.** Rather than heavily emphasizing one sub-stage, the best-performing portfolios consistently allocated capital across Pre-Seed, Seed Inception, and Seed Expansion. This balanced approach capitalizes on varied entry points along a startup's lifecycle while still preserving enough coverage at each stage to detect potential "super-performers."
- **Fund Size Matters Less Than Discipline.** Although the optimization occasionally gravitated toward larger fund sizes, the observed differences across the $110M–$150M range were far smaller than the impact of a sound capital allocation scheme. Once a fund is large enough to deploy in multiple categories with modest follow-ons, further increases in total capital showed diminishing returns—unless the fund's sourcing quality can also rise proportionally.
- **Maintain a Flexible Discovery Pool.** Although not as critical as limiting follow-ons and spreading initial checks across multiple stages, our results suggest that a modest discovery investment bucket can improve portfolio adaptability. By earmarking around 5–20% of the fund for opportunistic initial investments, managers can capture unforeseen high-upside deals without compromising the core early-stage focus. Over-allocating to discovery investment, however, can diminish the effect of a disciplined stage strategy and reduce total DPI.

In short, once a reasonable degree of diversification and stage coverage is achieved, *how* the capital is allocated (particularly in limiting follow-ons) generally exerts a stronger influence on DPI than a *small* up- or down-sizing in overall fund capacity. Funds that systematically maintain broad early exposure and restrain follow-on reserves are best positioned to capture outsized returns in a power-law environment.

8.1 Limitations

Although our enhanced simulations reinforce the broader notion that even spreads and low follow-ons drive performance, certain caveats remain:

- **Sourcing Constraints.** We assume the fund can consistently source enough qualified opportunities at each stage to deploy allocated capital effectively. If suitable deals at Pre-Seed or Seed Expansion are sparse, a balanced portfolio may be difficult to realize in practice.
- **Homogeneity of Exit Distributions.** The power-law exit model is held constant across all deal categories. In reality, deeply technical or capital-intensive sectors might exhibit different return dynamics than, say, consumer-software segments. Tailoring the distribution parameters by domain could reveal more stage-specific nuances.
- **Static Allocation vs. True Dynamism.** Real-world GPs adapt allocations mid-fund in response to performance signals, market conditions, or macro shocks. Our model sets allocations *ex ante*, which, while informative, omits any adaptive re-balancing that could further improve results.

- **No Fund-as-a-Whole Synergies.** We do not model the possibility that success in one deal might confer advantages (e.g., network, follow-on connections) that improve the chance or quality of subsequent investments. In reality, strong early deals can enhance brand and deal flow, potentially impacting future round access.

Despite these limitations, the repeated simulation and Bayesian optimization approach still offers useful guidance by highlighting key levers (especially the importance of broad initial coverage and limited follow-on reserves) that are likely to hold in a variety of real-world scenarios. Future extensions might include sector-specific calibrations or dynamic fund re-balancing strategies to capture second-order effects.

8.2 Future Work and Practical Extensions

Looking ahead, several possible extensions could strengthen the realism and utility of this framework:

- **Dynamic Rebalancing:** Introducing a re-optimization step after certain deal outcomes are realized, thus simulating how GPs can pivot strategy mid-fund.
- **Sector-Specific Tuning:** Adjusting failure rates, valuation uplifts, and power-law exponents for distinct domains (e.g., Security, Infrastructure & Tooling vs. Gaming, Community & Leisure).
- **Additional Performance Metrics:** While DPI was our chosen target, exploring optimizing IRR as a parallel objective might yield interesting results.
- **Multi-Fund Cohesion:** Coordinating multiple funds or parallel strategies (e.g., a dedicated "opportunity fund" for Series B/C investments).

We hope this study illustrates the benefits of *simulation + optimization* in demystifying VC fund construction. By systematically enumerating possible strategies in a realistic environment, GPs can better tailor capital allocation to their risk/return profile, deal-flow expectations, and LP preferences.

8.3 Final Reflection

Our initial motivation in building this simulation-optimization infrastructure was to identify an "optimal" portfolio construction across Pre-Seed, Seed Inception, and Seed Expansion. While we did uncover a set of strategies that outperform others on average, particularly those that limit follow-ons to a modest range and distribute capital in a balanced way among the different ticket categories, our process revealed that once these broad guidelines are satisfied, *further optimization has minimal effect.*

In short, many allocations within the same general range (e.g. whether Pre-Seed is 17% vs. 20% or whether follow-ons are capped at 12% vs. 15%) end up producing nearly identical average DPI. This suggests that, rather than from meticulous tuning of each percentage point, effective portfolio construction should focus on getting the *core principles* right: pursuing ample coverage to catch outlier successes, keeping follow-ons at a minimal level, and maintaining some flexibility with discovery investments. Beyond that, further effort should be spent on optimizing initial investment decision quality.

A Appendix: Optimization Results & Analysis

In this section, we present an expanded set of results from the Bayesian optimization routine described in Section 6. After running over 300 trials (subject to stopping criteria) with a parallelized Monte Carlo simulation of 10,000 samples (e.g. full fund simulation with a given portfolio construction) for each trial, we evaluated how the allocation parameters ($\theta_{\text{pre}}, \theta_{\text{inc}}, \theta_{\text{exp}}, \theta_{\text{co}}, \rho$, and a discrete choice of fund sizes) influenced the expected DPI of the fund (Distributions to Paid-In).

A.1 Identified Optimal Allocations

Through the final stage of the optimization, the best parameter vector discovered yielded the following:

- Pre-Seed: Initial 37.01%
- Seed Inception: Initial 24.04%
- Seed Expansion: Initial 18.15%
- Discovery: Initial 16.62%

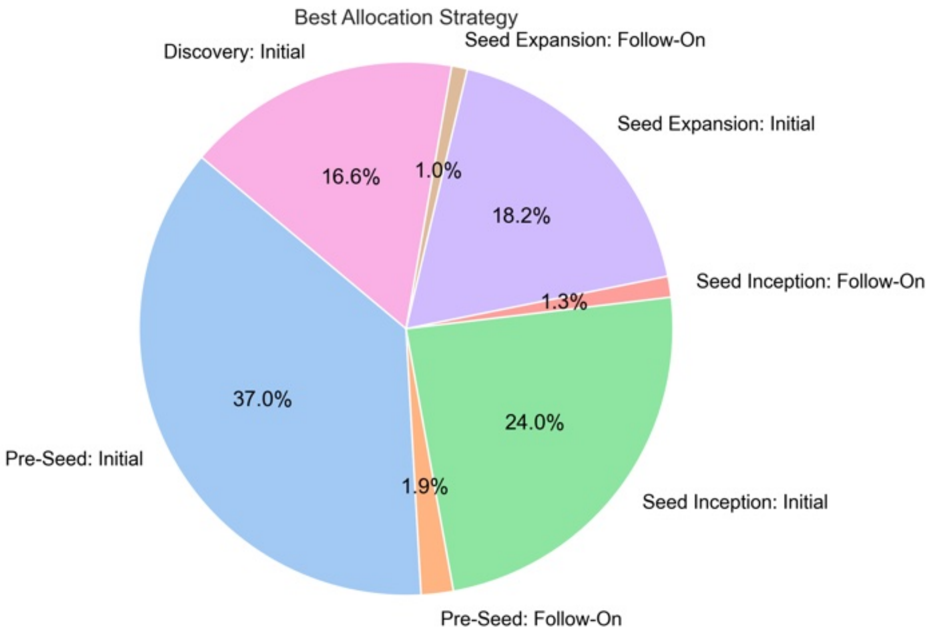

Figure 1: A pie chart showing how the investable capital was spread across ticket categories in the optimal portfolio construction configuration.

One of the most salient observations is the importance of *balanced diversification* across Pre-Seed, Seed Inception, and Seed Expansion stages. Historically, many VC funds focus heavily on a single sweet spot (e.g., purely at Seed or purely at Series A). Our results suggest that the distribution of exposure across different early-stage entry points, and *maintaining minimal but non-zero reserve* for each stage's follow-on commitments, produces higher expected DPI. We believe this emerges directly from the power-law return distribution: capturing a broader funnel of deals raises the odds of hitting a mega-winner, while still having sufficient follow-on resources to preserve ownership in that outlier.

A.2 Allocation Distributions and Exploration

Figure 2 shows a violin plot capturing the distribution of the *final* 7 allocation fractions (Pre-Seed Initial, Pre-Seed Follow-On, Seed Inception Initial, Seed Inception Follow-On, Seed Expansion Initial, Seed Expansion Follow-On, and Discovery Initial) sampled throughout the Bayesian optimization search. We observe a broad coverage of the parameter space, an expected outcome given the large number of Monte Carlo trials and the adaptive exploration procedure.

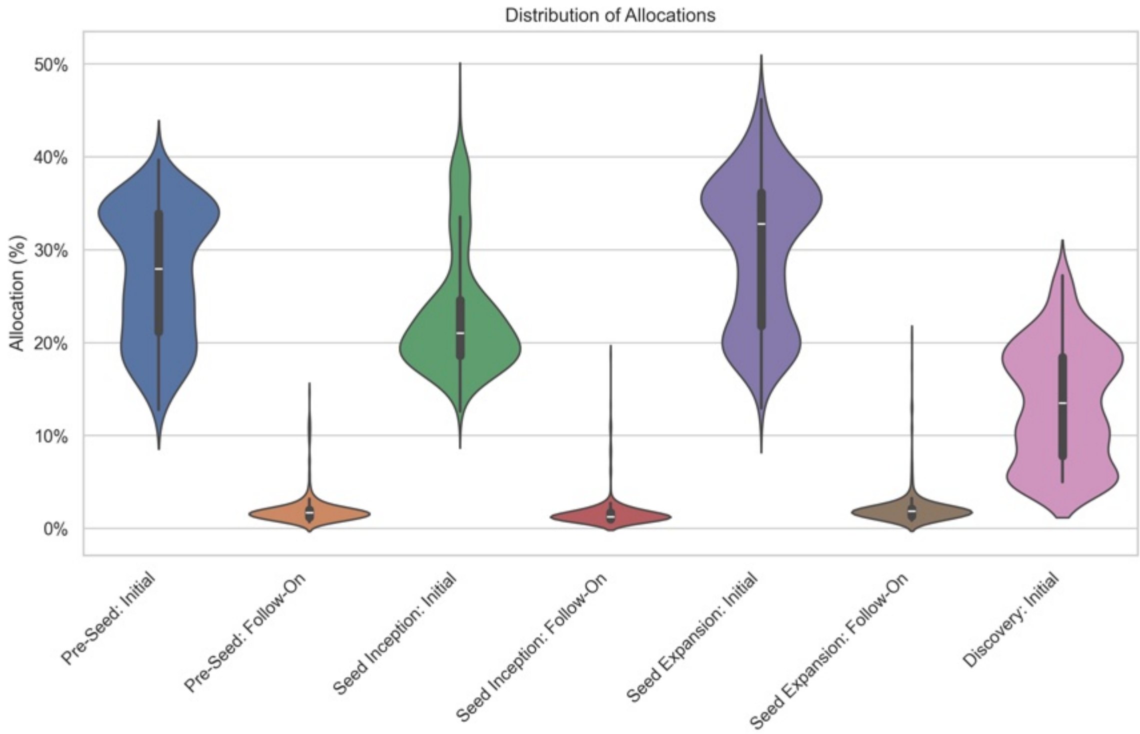

Figure 2: Violin plot of the 7 final allocation fractions across all trials. Each fraction is shown on the vertical axis (ranging from near 0% to around 40%), illustrating the range of allocations the optimization explored.

Based on these violin plots, we see that the initial investment allocation for each category has a relatively wide range of acceptable values while the follow-on allocations consistently choose much lower values within their allowable range. This suggests that the optimization process prioritizes flexibility in the initial capital deployment to explore opportunities across a broad spectrum of possible investments, while follow-on allocations are constrained by stricter criteria. The narrower range of follow-on allocations potentially reflects a more cautious approach, ensuring that subsequent investments are made with greater certainty and informed by the performance of prior allocations. Such a pattern aligns with a conservative risk management strategy, where early-stage investments are exploratory, and follow-on investments are reserved for opportunities with clearer potential for growth or return.

This behavior underscores the balance between exploration and exploitation inherent in the optimization process. By allocating more broadly in the initial stages, the model maximizes the opportunity to identify high-potential investments, while the tighter follow-on distributions ensure capital is concentrated in the most promising areas, optimizing overall portfolio performance.

A.3 Fund Size Variation

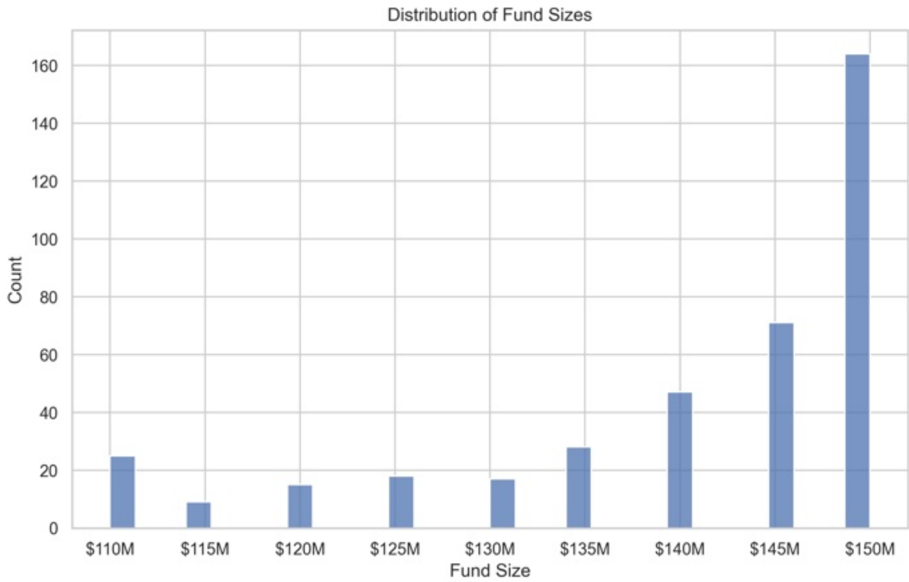

Figure 3: Histogram of fund sizes explored in the updated optimization runs. While there is a significant cluster around $150M, other executions of the optimization produced more erratic results.

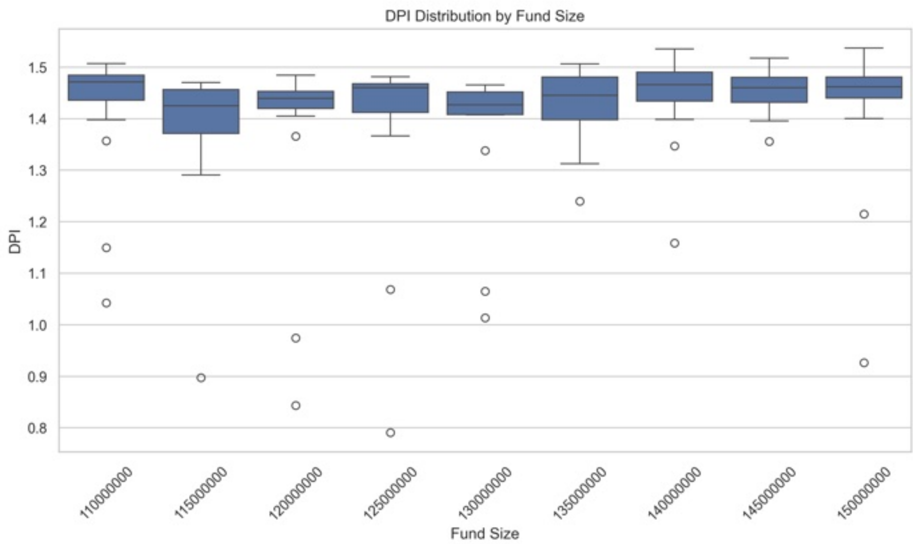

Figure 4: Box plot of fund sizes compared to DPI results. As we can see, despite the optimizer having a anecdotal preference for larger fund sizes, there is not a strong correlation between fund size and DPI.

Our parameter space included a discrete range of fund sizes from $110M to $150M in $5M increments. Figure 3 illustrates how often the optimizer sampled each of these sizes across the new set of runs. In an effort to test the durability of the optimization process, we ran multiple independent executions of the optimization. In all cases, one of two situations emerged: either the optimizer preferred larger fund sizes or the optimizer sampled fund sizes relatively randomly. Yet, in the final trial that we present here, the single *best* allocation strategy (in terms of DPI) emerged at $150M.

This pattern underscores two important points. First, while the optimizer can hone in on a particular fund size that best balances diversification and follow-on capacity in any given simulation, repeated runs show that the overall differences in expected DPI among fund sizes are relatively modest. Second, small variations in assumptions - such as deal availability, allocation fractions, or power-law tail behavior - can cause the algorithm to favor slightly different fund sizes from run to run. Rather than indicating a strong preference for larger funds, this suggests that the optimizer is responding to the nuances of each simulation instead of converging on a single "optimal fund size". This reinforces the broader takeaway that *how* capital is allocated (i.e., among Pre-Seed, Seed Inception, Seed Expansion, and Discovery) often proves more influential than simply whether the fund is $120M, $125M, or $150M in total.

From a practical standpoint, these findings suggest that a fund manager's primary focus should remain on crafting a well-structured allocation strategy—e.g., balancing initial coverage versus follow-ons—rather than fixating on adding or removing $5M–$10M of committed capital. Provided the fund can source enough high-quality deals to remain fully deployed, shifting among mid-sized fund targets (in the $110M–$150M range) likely exerts a smaller impact on expected DPI than does managing the underlying investment policy itself.

A.4 Sensitivity to Follow-On Allocations

A notable trend emerges when we examine *slice plots* of the simulation outputs, focusing specifically on how changes in the fraction of capital allocated to follow-on rounds affect portfolio outcomes. In these plots, we fix other parameters (e.g., the relative allocations to Pre-Seed, Seed Inception, and Seed Expansion) and then sweep the follow-on fraction ρ from lower to higher levels. Figure 5 provides one such example, illustrating how DPI correlates with ρ.

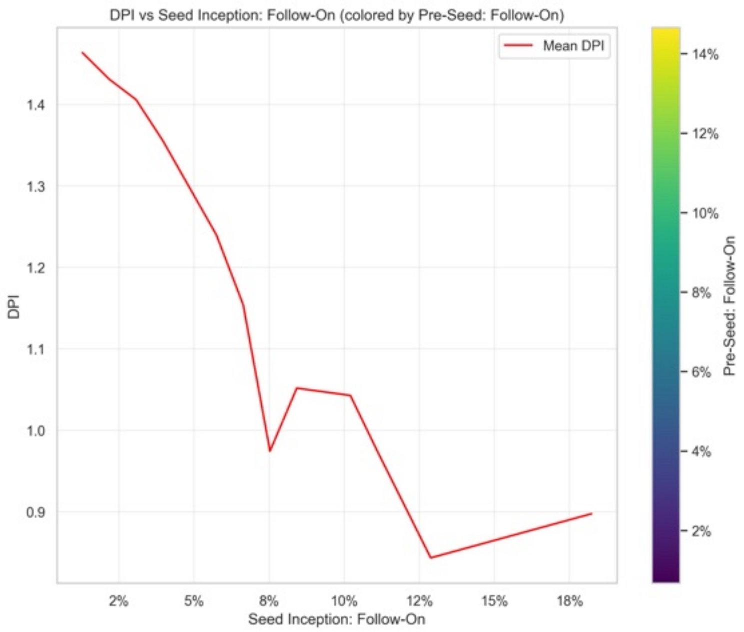

Figure 5: Slice plot showing how DPI trends as the follow-on fraction ρ increases. Each point represents a distinct simulation run, with other parameters fixed. The downward slope indicates an overall decline in expected performance as more capital is reserved for follow-ons.

We observe that DPI declines *quite predictably* as follow-on allocations approach the higher end of the range, suggesting that diverting an excessive portion of the fund to post-initial checks may undermine returns. Two core effects appear to drive this phenomenon:

1. **Reduced Initial Coverage:** Increasing the follow-on fraction shrinks the initial deployment pool. Because power-law outcomes hinge on a few outliers, excessively limiting the number of initial positions decreases the probability of capturing those rare "mega-winners."

2. **Capital Lock-In:** Although follow-on reserves can secure pro rata in successful companies, the high failure rate in early-stage VC means a large portion of that reserved capital may end up either unused or invested in "middle-of-the-road" follow-ons that do not offset the broader opportunity cost.

Put differently, while some follow-on reserve may be beneficial for defending ownership in obvious winners, our simulations indicate that a *moderate* follow-on fraction consistently outperforms more extreme policies. This aligns with the broader conclusion (Section A) that optimal strategies typically allocate only a modest proportion of each category's budget to late-stage checks, thus striking a balance between *breadth of initial bets* and *preserving enough ownership* in truly exceptional deals.

A.5 DPI Improvement Curve

A hallmark of successful Bayesian optimization is the progressive improvement in the objective metric as trials proceed. We tracked the cumulative maximum DPI found after each trial (see Figure 6). Early in the search, random or less-informed allocations produce baseline average DPI values in the 0.8–1.4 range. Over thousands of iterations, better parameter combinations are discovered, pushing the best-found average DPI close to 1.6.

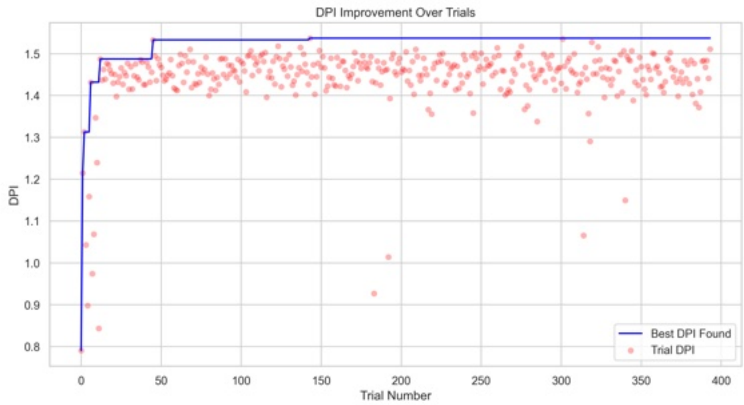

Figure 6: Cumulative maximum of DPI over trial index. The red scatter points represent actual trial DPIs, while the blue line is the running maximum. The plateau suggests the optimizer has found a near-stationary optimum.

Typically, during Bayesian optimization processes, we expect the performance of each trial to trend upward as the algorithm increasingly focuses on effective parameter spaces. However, in this case, the trial-by-trial DPI values exhibit significant variability throughout the search, with no clear upward trend. This suggests that the optimization landscape may be complex, with a combination of local optima and noise influencing the observed trial outcomes. Despite this, the running maximum DPI steadily increases, indicating that the optimization algorithm is effectively navigating the search space and identifying progressively better parameter combinations over time.

The observed variability in individual trial outcomes may also highlight the role of stochastic elements in the process, such as the inherent randomness in Monte Carlo simulations or the interplay of highly sensitive parameters. This reinforces the importance of viewing cumulative metrics, such as the running maximum DPI, to assess overall performance improvements rather than relying solely on individual trial results. Additionally, the plateau observed in the running maximum suggests that the optimizer has converged on a near-stationary optimum, offering confidence in the stability and reliability of the identified solution. This behavior underscores the utility of Bayesian optimization in uncovering high-performing configurations, even in scenarios characterized by noisy or non-monotonic response surfaces.

Another notable aspect of this optimization process is the implication of the plateau in the running maximum DPI for practical decision-making. The leveling-off observed in Figure 6 suggests diminishing returns from further exploration, a common feature of well-tuned Bayesian optimization procedures. This may be interpreted as a signal to transition from the exploration phase to deploying the optimized parameters in real-world scenarios. Moreover, the ability to identify this plateau can help reduce computational overhead by halting the search once a satisfactory solution has been reached, saving both time and resources. This underscores the importance of monitoring cumulative trends not only as a diagnostic tool for algorithm performance but also as a practical guide for determining when to conclude the search.

A.6 Detailed Allocation–DPI Relationship

To understand how each allocation fraction correlates with the DPI of the funds, we generate scatterplots of each fraction versus the resulting DPI (Figure 7).

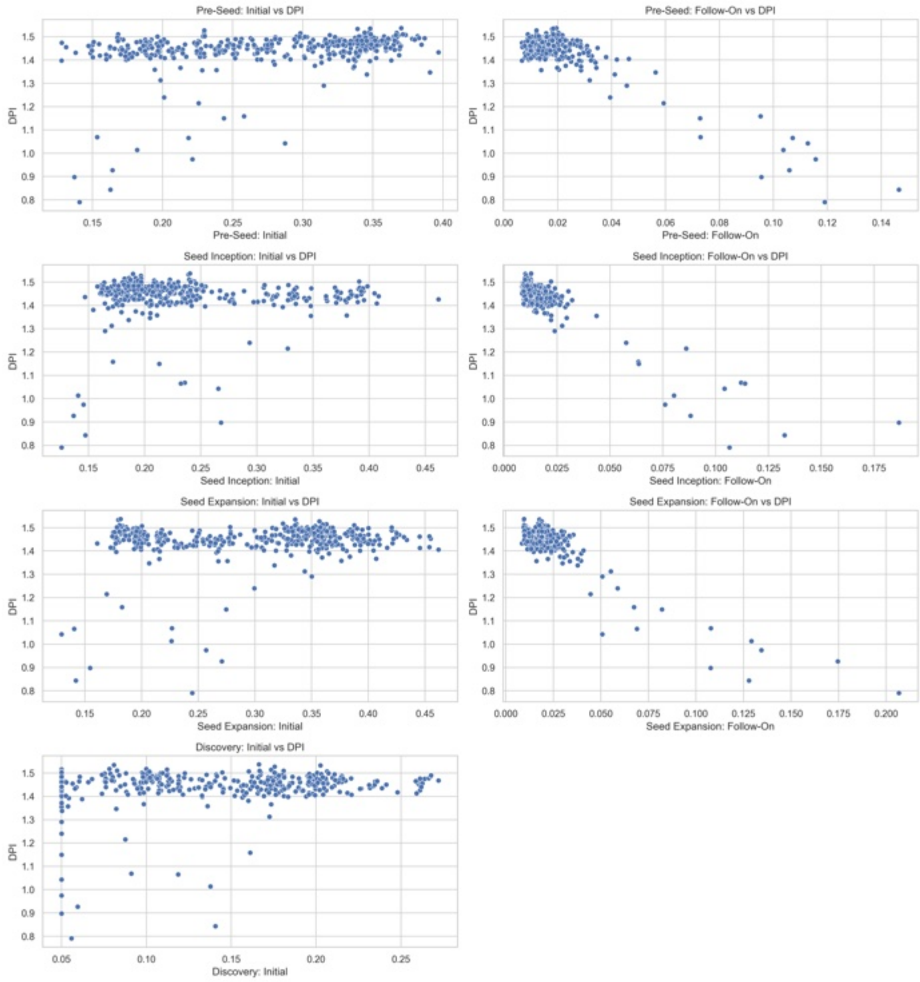

Figure 7: Scatterplots of each major category fraction (x-axis) vs. DPI (y-axis). Clustering regions indicate where certain allocations systematically outperform others.

The scatterplots reveal a recurring pattern for follow-on allocations, with higher DPI outcomes clustering at the lower end of the follow-on fraction scale. These observations reinforce the notion, first highlighted by our slice plots, that while some follow-on reserve is necessary to support breakout winners, overspending on follow-ons can dilute overall performance by reducing the breadth of initial bets.

B Appendix: Implementation Architecture

This section discusses how our conceptual framework for simulating and evaluating a venture capital fund is laid out in software. Although the code is not presented as a whole, we provide a high-level overview of the architecture, highlighting how different modules interact, the data they manage, and the sequence in which events occur. Our goal is to demonstrate the cohesive design that translates parameters and user configurations into a full Monte Carlo simulation of deal flow, follow-on investments, and exits.

B.1 Overview

Our methodology constructs a synthetic venture portfolio by:

1. **Allocating a Fund:** Defining the total investable capital (after fees) and splitting it across multiple investment categories.
2. **Scheduling Deals:** Generating initial deals within an investment period, grouping them into quarterly windows for capital calls.
3. **Tracking Stage Progression:** For each company, applying sequential rounds of *failure checks*, *valuation uplifts*, and *dilution events*, optionally factoring in signaling risk.
4. **Final Exit:** Sampling a holding period from a Beta distribution, then drawing an exit multiple from a truncated power-law distribution. Computing realized proceeds for the investor, which sum to the fund's distribution.
5. **Computing Metrics:** Calculating IRR, DPI, TVPI, and MOIC at the portfolio level, optionally repeating the entire simulation many times to estimate outcome distributions.

In the material that follows, we will explore how these conceptual steps are encoded in software, illustrating the design patterns and object model that make the simulation extensible and modifiable for real-world scenarios.

B.2 Core Components and Data Flows

Our implementation consists of three main components:

1. **Configuration Component:** Classes and objects specifying the fund's parameters, category definitions, and distributions (e.g., power law, beta) for exit timings or follow-ons.
2. **Portfolio Construction Component:** Functions and logic that transform these configurations into *actionable policies* (e.g., how many deals to generate, how large checks should be, and how capital calls are scheduled).
3. **Runtime Execution Component:** The simulation loop itself, where deals are created, positions are updated through multiple stages, and final exits are computed. This component also accumulates metrics like IRR, DPI, and so forth.

Although each component operates on distinct classes (e.g., `FundConfig` vs. `Portfolio`), they form a pipeline: configurations feed the construction logic, which in turn informs the simulation loop, which finally returns aggregated performance metrics.

B.3 Configuration Component

Fund-Level Settings. Users begin by specifying a *fund size*, a management fee rate, and an investment period in years. The `FundConfig` class validates these inputs, ensuring, for instance, that the investable capital (fund size minus total fees) remains positive.

Category Definitions. In parallel, a `CategoryConfig` is provided for each conceptual investment bucket (e.g., "Pre-Seed," "Seed Inception," "Discovery"). Each category includes parameters for:

- The *initial* investment policy: typical check sizes, minimum/maximum ownership, and any stage-specific details.
- Optional *follow-on* policies dictating how much additional capital is reserved for subsequent rounds (Series A, Series B, etc.), and under what strategy (pro rata vs. full ticket).

A user might define multiple categories if they plan to diversify across various round sizes or investment themes.

Distribution Choices. Finally, a `PowerLawConfig`, `FollowOnConfig`, and `ExitConfig` define the random draws for:

- **Exit Multiples:** Governed by a truncated power-law distribution (Section 4), ensuring that rare but large outcomes can occur.
- **Follow-On Timing:** Modeled with a Beta distribution, translating uniform samples into a realistic spread of months or years after initial investment.
- **Exit Timeline:** Another Beta distribution for how many years (post-investment) a surviving company waits before exiting.

By encapsulating these in separate configuration classes, we enable users to override individual assumptions (e.g., a heavier power-law tail or a faster typical holding period) without altering the rest of the system.

B.4 Portfolio Construction Component

Once the user defines a complete `SimulationConfig` (gathering `FundConfig`, category configurations, and distribution parameters), the code calls a *portfolio construction* function. This logic:

1. **Computes Investable Capital.** Subtracting total management fees from the fund size yields the net capital actually available for deals.
2. **Allocates Capital by Category.** Each `CategoryConfig` specifies an *allocation fraction*; these fractions must sum (approximately) to 1.0. The code multiplies each fraction by the total investable capital to determine each category's "budget".
3. **Sets Initial and Follow-On Ticket Ranges.** For each category, the *initial* stage has a recommended ticket size policy (e.g., \$500k–\$1.5M). Similarly, each follow-on policy lists min/max check sizes for subsequent rounds.
4. **Computes Expected Deal Counts.** Given each category's budget and average ticket size (estimated from the chosen range), the system infers how many deals the category can attempt. This is only an approximate count, since random draws may yield slightly different ticket sizes in practice.

The output of this process is a `PortfolioConstruction` object that effectively maps user-defined policy inputs into actionable instructions for the simulation loop (i.e., "we aim for N deals in Category A, each with tickets in [\$X, \$Y]," and so on).

B.5 Runtime Execution Component

1) Deal Generation. Armed with the `PortfolioConstruction` specifications, the simulation creates a `Portfolio` object. It samples random *inter-arrival times* for initial deals (using a Poisson-like or exponential distribution), so that the deals are spread over the investment period. For each initial deal, the system also probabilistically schedules one or more *follow-on* deals, spaced according to the Beta distribution from the `FollowOnConfig`. These deals are sorted by date and grouped into calendar "quarters" to mimic periodic capital calls in real life.

2) Capital Calls and Execution. At each quarter boundary, the simulation totals the capital needed for transactions occurring within that quarter. A *capital call* is made for that amount (or a round approximation) as long as uncalled investable capital remains. If the fund still has capacity, each scheduled deal is "executed":

- **Initial Deal:** Creates a new `Position` in the `Portfolio`, recording ownership fraction, ticket size, and date.
- **Follow-On Deal:** Updates an existing `Position` to reflect additional shares purchased, changes in stage valuation, and resulting dilution for non-participating investors.

If capital is insufficient, some deals may be skipped, capturing the reality that funds can run out of deployable reserves before the end of the investment period.

3) Stage Survival and Signaling. After all deals are placed, each `Position` is retrospectively checked for survival through its invested stages. For example, if a fund invests at the Pre-Seed round and again at the Seed round, it calculates the probability of failing before Series A. The `FollowOnConfig` enforces a signalling risk: if a large investor chooses not to re-invest at the next stage, the chance of failure for that startup can increase. Positions that fail are written off at $0.

4) Exit Events. Surviving positions eventually exit at a random time drawn from the Beta distribution in `ExitConfig`. The multiple on invested capital is sampled from the truncated power-law distribution. Multiplying the total invested amount by that multiple yields the final return, which the simulation records in the `Portfolio` as a realized distribution (thereby increasing Distributions to Paid-In, etc.).

B.6 Summary of the Implementation

In summary, the implementation of our system can be understood as a pipeline:

1. **Configuration and Validation:** The user sets up `FundConfig`, `CategoryConfig`, and distribution parameters. We ensure that the allocations sum to a coherent value and that the ticket sizes, ownership targets, and stage definitions are consistent.
2. **Portfolio Construction:** A specialized function translates these parameters into budgets, deal counts, and ticket-size policies. It also initializes the fundamental categories that will guide investment decisions.
3. **Runtime Execution:** The simulation loop:
    - Generates deals over time (initial vs. follow-on).
    - Issues capital calls quarterly, invests in deals if capital remains.
    - Applies stage-wise failure checks and updates valuations.
    - Concludes with an exit event for every surviving position, using Beta timing and a power-law multiple.
4. **Performance Calculation:** IRR, DPI, TVPI, and MOIC are computed at the end of each run, and aggregated across multiple Monte Carlo runs to generate statistical profiles of the outcomes.

This architecture provides a clear separation between policy definitions (the "what" of investment strategy) and runtime mechanics (the "how" of capital deployment and exit realization). By maintaining consistent interfaces among these components, we ensure that changing any single dimension (e.g., altering the power-law exponent or adjusting category allocations) remains straightforward.

C Appendix: Beta Distribution Parameterization

A key component in our simulation logic is the Beta distribution, which we use to draw holding periods for follow-on investments and exit events (see Section 5.5). Mathematically, the Beta distribution is defined for random variables $X \in [0, 1]$, and is parameterized by two positive real numbers $\alpha > 0$ and $\beta > 0$. Its probability density function (PDF) is given by

$$f_{\text{Beta}}(x;\, \alpha, \beta) \;=\; \frac{x^{\alpha-1}\,(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)} \quad \text{for } x \in [0, 1], \tag{1}$$

where

$$\mathrm{B}(\alpha, \beta) \;=\; \int_0^1 t^{\alpha-1}\,(1-t)^{\beta-1}\, dt$$

is the Beta function (a normalization constant ensuring the PDF integrates to 1).

Interpreting α and β.

- When $\alpha < 1$ and $\beta < 1$, the density has peaks near $x = 0$ and $x = 1$, pushing draws to the extremes.
- When $\alpha > 1$ and $\beta > 1$, the density is more "bell shaped," centered in the interior of $[0, 1]$.
- Larger α (relative to β) skews the distribution toward 1, whereas larger β (relative to α) skews it toward 0.

In our context, x represents a normalized fraction that is then scaled onto $[t_{\min}, t_{\max}]$—for instance, 4 to 10 years. The variation of α and β therefore changes whether the exit times tend to be "front loaded" or "back loaded".

Illustrative Parameterizations. Figure 8 shows six different Beta distributions, each with a distinct (α, β) pair. Note how changing α and β affects the shape and location of the PDF over $[0, 1]$. In practice, we pick values that match observed patterns in average holding periods and exit timing, but these parameters are user-configurable to model different market conditions.

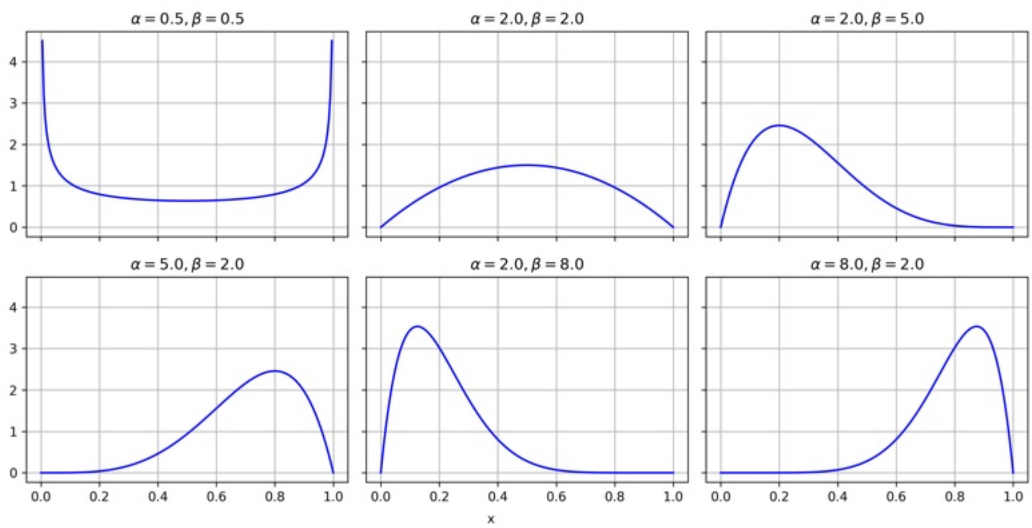

Figure 8: Examples of Beta distributions for different (α, β) parameter values. Shifting α and β modifies the skew or central mass of the density over $[0, 1]$.