Experimentation in Complex Environments with Learning Spillovers: Applications to New Product Development*

Yuqi Yang† Harish Guda‡ Hongmin Li§

Abstract

Problem Definition: In new product development, firms/managers experiment amongst a collection of alternatives before finally choosing a particular product (design choice) for mass production (commercialization). Experimentation in such environments involve *learning spillovers*: firms learn not only about the product they test, but also about similar choices that are untested. Our model captures three practically relevant features in such environments: a vast plethora of feasible products, complex product-performance mappings where small design changes may result in large differences in outcomes (local unlearnability), and learning spillovers that are stronger when products are more similar (partial invertibility). We analyze the firm's experimentation strategy and its implications for commercialization.

Methodology: We consider the set of products as the unit interval, and the product-performance mapping as a realization of a Gaussian process. We cast the firm's experimentation problem as a dynamic Bayesian optimization problem.

Results: We present an analytical solution to the firm's problem, i.e., we identify an optimal experimentation policy for the firm (the product to be searched in each period, and the product that is mass-produced upon concluding experimentation). Motivated by computational challenges imposed by the optimal policy, we present a structurally consistent and computationally easier heuristic that we refer to as segment-based search. We demonstrate the performance of both these policies relative to a non-adaptive benchmark.

Implications: From a substantive standpoint, our analysis makes a fundamental contribution in presenting an analytical approach to solving the firm's experimentation problem. The firm's objective in our model is identical to the "expected improvement" problem in Bayesian optimization, where the extant literature has primarily resorted to numerical approaches to solving this problem. From a managerial standpoint, we provide a prescriptive approach to how firm's should experiment when they are faced with a continuum of products. We also demonstrate the pitfalls that arise under both proposed policies: the optimal policy under-invests in experimentation if it identifies early-winners, while the segment-based search over-invests in experimentation, even in regions of low potential.

Keywords: New Product Development, Learning Spillovers, Gaussian Process.

*Latest version available here.

†W.P. Carey School of Business, Arizona State University. yyang450@asu.edu
‡W.P. Carey School of Business, Arizona State University. hguda@asu.edu
§W.P. Carey School of Business, Arizona State University. hongmin.li@asu.edu

1 Introduction

In many business decision-making problems, managers experiment amongst a set of viable alternatives in their search for the best choice. Examples of such settings relevant to Operations Management (OM) include a broad spectrum of decisions in new product development. For concreteness, consider the following examples:

- In the apparel industry, merchandising managers experiment across colors in their search to identify the optimal color (intensity) for fashion products;
- In clinical drug development, pharmacologists experiment with dosage levels in their search to identify the optimal proportion of ingredients, or dosage of a medication;
- In the food and beverage industry, packaging engineers and food scientists experiment with both packaging and flavors in their design of confectionary products;
- In UX/UI design, product managers experiment with layout/design of webpages/apps, in their search for the most intuitive/appealing design.

Such search processes often involve *learning spillovers*: firms learn not only about the design choice they test, but also about similar choices that are untested. This is because products "in the vicinity" of the tested product are unlikely to have vastly different outcomes. Our work in this paper is motivated by decision-making in environments that involve learning spillovers.

1.1 Illustrative Example

Our motivating example in this paper draws on the process of experimentation at Shein, a global e-commerce platform specializing in fast fashion (The Wall Street Journal, 2023).[1] Consider the product line for t-shirts of different shades of gray, as shown in Figure 1. Each color corresponds to a unique product.

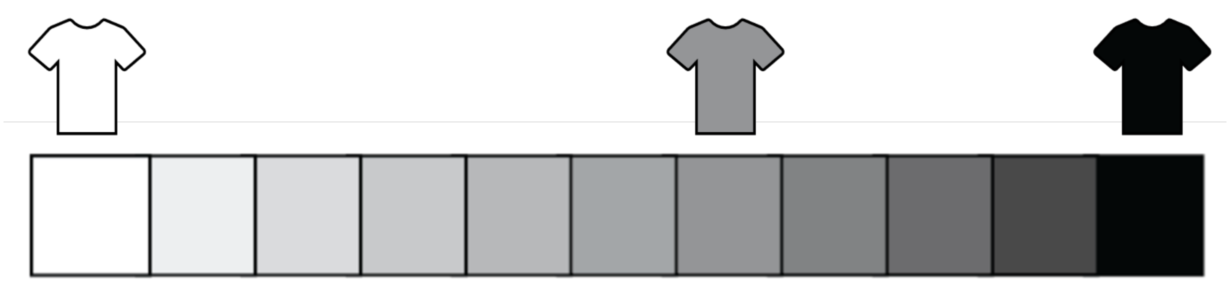

Figure 1. Motivating Example: Experimentation for a Merchandising Manager

[1] See https://us.shein.com/

Assume that the firm sequentially tests t-shirts of different shades of gray, before it chooses a specific shade of gray to mass-produce. Suppose the firm has tested two products, say at the boundaries: the t-shirt in white (left boundary) and the one in black (right boundary). Consider a product in-between the boundaries, say the t-shirt in an intermediate shade of gray. Naturally, popularity of the white t-shirt at the left boundary may provide signals about the market potential of t-shirt with colors in its immediate vicinity (e.g., off-white, dove-gray, etc.); whereas the popularity of the back t-shirt at the right boundary signals the market potential of its immediate neighbors (e.g., near-black, midnight-gray, etc.). We refer to such information that tested products provide for untested products as learning spillovers.

In general, tested products provide feedback about both tested and untested products to merchandising managers. This feedback then guides their search for a better product. The performance of products in the vicinity act as informative signals of the performance of the new product. Further, the tested products are more informative of the new product if they are close to each other. As the distance between the tested and untested products grows, the extent of learning spillover dissipates and, consequently, the performance of the untested products becomes further dominated by the unknowns.

Below, we explain the features of such environments involving learning spillovers, and then provide a preview of our analytical model and main results.

1.2 Learning in a Complex Environment

Our work in this paper focuses on learning and decision-making in *complex* environments (Callander, 2011). A complex environments differs from a *simple* environment in several aspects. To explain these differences, we use Shein's problem of new product design/testing to learn consumer tastes as a running example of a complex environment throughout this section.

First, the plethora of products that Shein could choose from – their test-bed – is vast. Even for a simple gray t-shirt, there is an infinite number of shades to experiement with. Identifying the highest-performing choice under an accelerating fashion clockspeed requires efficient testing and rapid learning. In fact, designers at Shein develop new products/designs by borrowing knowledge gained from existing products.

Second, mappings between products and their performances are incredibly complex. That is, even small perturbations to, say the design of an existing product (e.g., the color, style, etc.), can affect its performance. It is impossible to predict, with certainty, how such perturbations can affect the performance of the product. Callander (2011) refers to this phenomenon as product mappings being "not locally learnable". For example, while one may contemplate that an off-white t-shirt may have a similar popularity as the white t-shirt of the same style, it is possible that a subtle tint of gray may dramatically alter the aesthetic value of the product to the target customer segment and leads to a sharp boost or decline of interest. While Shein's designers may leverage comparable products from the past for potential demand signals, it is impossible even for experienced category managers to predict demand accurately (Baardman et al., 2018; Kavadias and Ulrich, 2020).

Third, while it is impossible to predict the performance of new products with absolute certainty, the performance of products that are similar provide partially informative signals about the performance of new products. That is, we postulate that a high degree of correlation between product similarities and their performances. Callander (2011) refers to this property as "partial invertibility". Indeed, demand forecasting for many new products often relies on the assumption that "comparable products perform comparably well" (Baardman et al., 2018). an existing product's performance is more (resp., less) informative about a new product's performance if it is more (resp., less) similar. That is, the informational content (correlation in product performances) diminishes in the extent of dissimilarity between the products. This is also commonly assumed in demand forecasting. For example, in demand forecasting, the relative weights given to different comparables depend on how similar they are to a focal product.

Therefore, in new product development, the effects of both "partial invertibility" and "local unlearnability" are at play. On the one hand, the performance of a tested product may shed *some* information about products in its vicinity; on the other hand, little is known about untested products and small differences can trigger drastically different market reactions. In this paper, we present a model that captures such seemingly conflicting features of reality, and develop a prescriptive approach for how firms should experiment across new products before commercialization.

1.3 A Preview of Our Model

The extant literature in OM that studies new product development assumes that a firm (testing a new product) has access to a finite collection of new products (viable alternatives) whose market potential/outcomes are unknown. Further, they assume that learning from testing a product reveals little about untested products. Given the contexts that motivate our work, our work deviates from the extant literature in several important ways. We consider a case where the set of products available to the firm are infinite and organized on a continuum. Specifically, we let the set $\mathcal{X}_\infty = [0, 1]$ denote the set of all products. To model the complexity in the product-performance mapping, we let the performance $f(x)$, $x \in \mathcal{X}_\infty$, be a realization of a Gaussian Process (a random walk in continuous space) chosen by nature. By definition, a Gaussian process is continuous almost everywhere (Garnett, 2023). This provides a framework to capture not only the learning spillover via "partial invertibility" that is fundamental to our work but also the locally unlearnable feature of the market reality.

We assume that the firm has a de-facto/status-quo product (denoted by $x = 0$). Time is discrete; the firm experiments with one product in each period. We model two phases of decision-making for the firm: an experimentation phase where they sequentially test products (lasting T periods), and a mass-poduction phase where they mass-produce one product. During the experimentation phase, if the firm chooses to experiment a product $x \in \mathcal{X}_\infty$, we assume that it learns $f(x)$.[2] The firm then updates its belief on the realized path $f(\cdot)$, and then proceeds in its search for the next product. After the experimentation phase, the firm concludes search and chooses a product to mass-produce. In this

[2]This is referred to as *noise-less* learning (or learning from exact observations) in Bayesian optimization; see Chapter 10 in Garnett (2023).

paper, we analyze the firm's experimentation policy (i.e., identify the product searched in each period during the experimentation phase) to maximize the profit from mass-production. We remark that our modeling approach involving two distinct phases – experimentation and mass-production – is an abstraction of reality. In practice, experimentation in new-product development is an ongoing process that may continue even after mass-production.

1.4 A Preview of Our Main Results

From a methodological standpoint, our work draws on the literature on Bayesian optimization. Our analysis contributes to this literature by providing analytical results that guide the firm's experimentation policy (i.e., the product experimented each period). During the experimentation phase, we express the firm's objective as a direct sum of two components: the best observed product in its history, and the value from continued experimentation. The value from continued experimentation is equivalent to "expected improvement", a widely employed criterion in the Bayesian optimization literature. However, much of the field has resorted to numerical approaches in their search for an optimal solution due to the objective function being neither convex nor concave (Garnett, 2023). In contrast, we provide an analytical characterization that identifies a unique optimal product to experiment with in each period.

Even though the optimization problem results in a unique solution, implementing the optimal experimentation policy is challenging if the horizon of experimentation is long, as the decision rule requires solving a sequence of optimization problems. Consequently, the computational burden accumulates over time. Motivated by a novel analytical structure in the expected gain from continued experimentation, we propose a simple heuristic, that we refer to as "segment-based search" that preserves certain desirable properties (e.g., the uniqueness of the optimal product to be experimented with in each period) while substantially reducing the computational burden as the search proceeds.

Finally, in Section 8, we numerically compare the performance of the optimal experimentation policy, the segment based search and a non-adaptive benchmark in terms of their simple regret (Bubeck et al., 2009) and the worst-case performance. We demonstrate that the optimal experimentation policy achieves a lower simple regret than the segment-based search policy, while both significantly outperform the non-adaptive benchmark. However, the segment-based search policy has a clear computational advantage over the optimal experimentation policy, and achieves a comparable worst-case performance as the optimal experimentation policy.

2 Literature Review

We review two streams of literature closely related to our work: the new product development literature in OM, and the experimentation and learning literature. Given the vast amount of work in both streams, we focus on work close to us.

2.1 New Product Development in Operations Management

The field of innovation and new product development spans multiple disciplines including Economics, Strategy, Marketing, and Operations Management. Below, we review papers that are close to our work. For a comprehensive review of the taxonomies on innovation and new product development in OM, we refer the reader to Kavadias and Ulrich (2020). Within the many strands of work in the new product development literature, our work focuses on the operational execution of new product development through iterative experimentation. In such environments, the firm sequentially tests prototypes, observes market responses, and iterate this process before commercialization. This practice is commonly referred to as "rapid prototyping" (Thomke, 1998).

Much of the analytical work in this stream focuses on the tradeoff between performance (of the mass-produced product) and the costs of experimentation (direct or opportunity costs, as experimentation is costly and takes time). In their seminal work, Weitzman (1979) analyzes how a decision-maker (firm) tests a collection of alternatives before terminating search. They assume that the performance of each product is independent and the firm incurs a fixed cost to experiment a product. They show that a remarkably simple rule – based on a comparison of reservation utilities of unsearched products and the performance of the best observed product in history – is optimal. Several papers build on this framework to understand the value vs. cost of experimentation, for example on the firm's test schedule or how many prototypes to develop; some examples from this stream include Thomke (1998) and Dahan and Mendelson (2001).

While much of the work in the above stream assume independence in the performance of new products, there is a nascent stream of work that has looked at the role of spillovers in new-product development. Erat and Kavadias (2008) study the effect of learning spillovers (referred to as transfer learning) on a firm's testing strategy. They assume that the set of products is finite, and the firm's belief about product performance is a multivariate normal distribution with equal variance. In addition to the (upper) threshold as in Weitzman (1979), they show the firm's decision to continue search depends on a lower threshold that reflects a low performing space (i.e., all products are likely low performers). Yoo et al. (2021) study the "lean start-up method" where a firm cycles through multiple build-, test-, and learn-phases, before the firm launches the final product or chooses to modify/pivot. In their model, true customer preferences are not observed; rather, firms learn through sales (a censored version of preferences). While the space of products is dense, they assume that the firm has only two opportunities to iterate (i.e., their model is a two-period model). In contrast, we solve the optimal experimentation policy for a horizon of any length $T \geq 2$ under a richer model of learning spillovers. Relatedly, Bhaskaran and Erat (2025) analyze the optimal amount of fidelity (or informativeness) in firms' design of new products, and how even low fidelity experiments can be optimal. More recently, Huh et al. (2025) consider a setting where a product's performance is the sum of a common component and an idiosyncratic component (random, product-specific), where inferring the common component allows for learning spillovers. Experimenting across products allows the firm to learn the common component in their search for the best product. They cast this problem as a Bayesian dynamic program, and propose an optimal two-threshold policy (akin to Erat and Kavadias

(2008)).

Our model departs from this literature in a few fundamental ways that are crucial for operational execution. First, the firm's set of feasible products form a continuum, so the action space is uncountable and infinite. Second, performance is correlated across products, and the extent of correlation is higher if the products are more similar. Third, we consider the firm's experimentation to last multiple periods (in particular, more than two periods, that many papers in the OM literature consider). Our work has much in common with Callander (2011), who pioneered the framework of using a realization of a Gaussian process (in their work, a Brownian motion with negative drift) to model product performance over a dense set of products. In their work, Callander (2011) assumes that the ideal product is the first zero of the Gaussian process, and adopts a quadratic loss form to model the firm's payoff.[3] Such an assumption significantly simplifies the firm's optimization problem, and is quite stylized and abstract. Finally, Malladi (2022) proposes a model of search with spillovers (referred to as spatial search), where the firm competes with an (imaginary) adversary in their search for the best product. They assume that the set of poducts is the unit interval. However, they assume that the product-performance mapping is either concave or quasi-concave. In contrast, we make minimal restrictions on $f(\cdot)$, and identify an optimal experimentation policy for the firm.

2.2 Experimentation and Learning

Within the literature on experimentation and learning, a decision-maker (the firm) makes adaptive sequential decisions in their search for the best among a collection of alternatives. Arguably, the most common modeling approach used is the multi-armed bandit (MAB) framework, where the set of alternatives is countably finite. In its simplest version, the optimal policy admits an index-based characterization, referred to as the Gittins index (Gittins, 1974). At its core, the decision-maker's balances exploration (of new arms) and exploitation (by choosing well-performing arms). There are a variety of algorithms proposed in the literature, with the most popular ones being the Upper Confidence Bound (UCB) algorithm (Auer, 2002), the ϵ-greedy algorithm (Auer et al., 2002), and Thompson Sampling (Thompson, 1933; Ferreira et al., 2018).

These ideas have been widely adopted in Operations Management for a variety of settings, including price optimization (Keskin and Zeevi, 2014; Broder and Rusmevichientong, 2012; den Boer and Keskin, 2022), assortment optimization (Agrawal et al., 2019), and inventory management (Besbes and Muharremoglu, 2013; Zhang et al., 2020). Across these applications, the objective is to develop algorithms that learn model primitives online and make optimal data-driven decisions that yield asymptotically optimal policies with provable performance guarantees. Frazier et al. (2009) extends this framework to a set of finite and correlated alternatives and proposes an optimal forward-looking policy. While this is analytically challenging, they demonstrate that the expected improvement policy can be close in performance to the optimal policy in this setting using numerical experiments. Ryzhov (2016) analyzes the general convergence of the expected improvement method under finite

[3] That is, in the context of our work, the firm's payoff from the mass production of x is of the form $-(f(x))^2$.

alternatives and derives the limiting allocation both under known and unknown variance.

Our setting lies outside the canonical scope of this literature in two dimensions that are both essential for experimentation and learning: infinite and highly correlated arms. As the number of alternatives (bandits) is infinite, the conditions of the optimal policy based on the Gittins index do not hold. As such, during each period of experimentation, the set of alternatives with the highest potential remains unknown, and solving the stochastic dynamic programming for an optimal policy is operationally infeasible due to the curse of dimensionality (Russo, 2016; Garnett, 2023). In our setting, the firm faces an *infinite* collection of designs indexed by a continuous design variable, and outcomes are *highly correlated* across designs. Each experiment, therefore, transfers information to a set of infinite untested designs through Bayesian updating, so the belief state is a posterior distribution over an unknown function on a continuous domain.

We contribute to this stream by providing a constructive resolution of the *best arm identification* problem in an infinite- and correlated-armed bandit problem. We show that the expected gain from continued experimentation admits a unique maximizer in every period. This result plays an important role analogous to the index structure in finite-arm models by pinning down a single optimal bandit for subsequent experimentation. This continuum of product choices not only enables richer qualitative insights but is also well suited to many practical decision contexts, as discussed in Section 1.

3 Model

Motivated by our applications for new product experimentation/development, we consider a firm that has access to a plethora of products that can be ordered in a sequence. Specifically, we assume that the set of (untested/new) products available to the firm is the unit interval $\mathcal{X}_\infty = [0, 1]$.[4] Each point $x \in \mathcal{X}_\infty$ corresponds to a unique product. We assume that the firm has a status-quo product, and is located at the left boundary of $\mathcal{X}_\infty$ (i.e., $x = 0$). For any two products $x_1, x_2 \in \mathcal{X}_\infty$, we define their *dissimilarity* by the distance $|x_1 - x_2|$.

Every product $x \in \mathcal{X}_\infty$ is associated with a performance index $f(x) \in \Re$, which captures the performance of the product along relevant economic dimensions (e.g., volume of sales, perceived quality, etc.). We refer to $f : \mathcal{X}_\infty \to \Re$ as the product-performance mapping. Recall from Section 1.2 that the product-performance mapping exhibits the following key properties:

(a) a high degree of uncertainty;

(b) involves learning spillovers: the performance of one product is informative of the performance of other similar products (in its vicinity).

(c) the informational content diminishes in the extent of dissimilarity between the products.

[4] In reality, products may not be mapped to a single dimensional space. Rather, they are represented in a multi-dimensional space. For such settings, we view our model as a building block.

(d) is not locally learnable: the mapping is not locally learnable if small differences in the product can (potentially) lead to large differences in the product performance;[5]

(e) is "partially" invertible: the mapping is not "fully" invertible in that it is impossible to fully identify $f(x)$ given any finite history (or observational data);[6]

To capture the complex relationship between the set of products and their associated performance, we model $f(x)$ as a realization of a Gaussian process with a Weiner kernel. For readers unfamiliar with a Gaussian process, we provide a brief description below.

Definition 1. *(Williams and Rasmussen, 2006) A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.*

A Gaussian process is characterized by two components – the mean function and the covariance function – that we explain below. Consider a (finite) collection of $k \in \mathbb{N}$ points, denoted by $\mathbf{x} = [x_1, \ldots, x_k]^\top$. Let $\mathbf{f}(\mathbf{x}) = [f(x_1), \ldots, f(x_k)]^\top$. Under a Gaussian process, $\mathbf{f}(\mathbf{x})$ follows the multivariate normal distribution

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{N}\left(\mathbf{m}(\mathbf{x}), \mathbf{\Sigma}(\mathbf{x})\right), \tag{3.1}$$

where

(i) the mean vector $\mathbf{f}(\mathbf{x})$ is defined via the mean function $m : \mathcal{X}_\infty \to \Re$, and

$$\mathbf{m}(\mathbf{x}) = [m(x_1), m(x_2), \ldots, m(x_k)]^\top;$$

(ii) the covariance matrix $\mathbf{\Sigma}(\mathbf{x})$ is defined via the Wiener kernel as follows. For any $x, x' \in \mathcal{X}_\infty$, define $k(x, x') = \sigma_0^2 \min\{x, x'\}$. Then,

$$\mathbf{\Sigma}(\mathbf{x}) = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_k) \\ \vdots & \ddots & \vdots \\ k(x_k, x_1) & \cdots & k(x_k, x_k) \end{bmatrix}.$$

Figure 2 provides an illustrative example of $f(x)$ with $m(x) = 0$ and $\sigma_0 = 1$. The Gausian process with Wiener kernel captures properties (a)-(e) that characterize the challenges in new product experimentation. As shown in Figure 2, on the one hand, $f(x)$ exhibits continuity which implies that tested products provide *some* information on the performance of untested products depending on the proximity; on the other hand, $f(x)$ can also shift drastically in a random-walk fashion at any given x, with the degree of such uncertain shift described by the parameter σ_0. As a result, this modeling choice of a generic product-performance mapping admits a wide variety of practical settings.

[5] Specifically, in our context, consider a product at x whose performance $f(x)$ is known, and an adjacent new product at $x + \epsilon$ for $\epsilon > 0$. We have that Supp $(f(x+\epsilon)|f(x)) = \Re$ and Var $[f(x+\epsilon)|f(x)] > 0$

[6] For example, if $f(x)$ is a polynomial function, say $f(x) = \sum_{k=0}^{K} \beta_k x^k$. Such a polynomial function is fully identified by any history of size $K + 1$

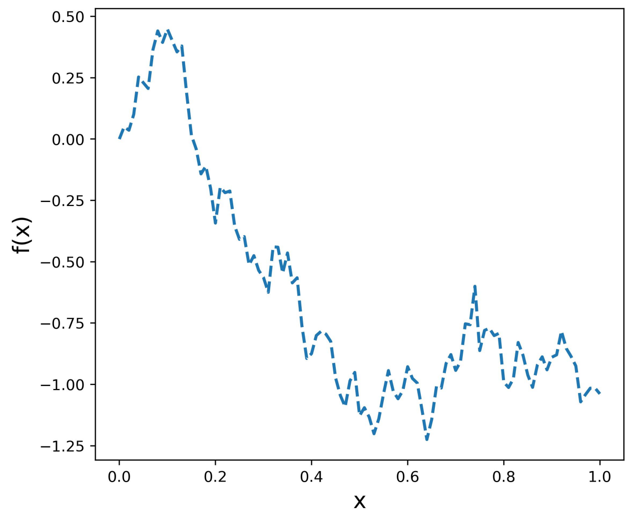

Figure 2. An Illustrative Example of $f(x)$, with $m(x) = 0$ and $\sigma_0 = 1$

Time is discrete and finite, indexed by $t = 1, 2, \ldots, T+1$. Let $[t] = \{1, 2, 3, \ldots, t\}$. There are two distinct phases in the firm's decision-making:

(a) Experimentation: The firm's experimentation phase lasts T periods. The firm experiments (tests) a product $x_t \in \mathcal{X}_\infty$ and learns its true performance $f(x_t)$. In particular, we assume that the firm learns $f(x_t)$ without any noise. At the start of $t = 1$, the firm only knows the performance of the status quo product, denoted as $f(0)$, which we normalize to 0. At the start of period $t \in [T]$, denote the set of products tested thus far by $\mathcal{X}_t$:

$$\mathcal{X}_t = \{0, x_1, x_2, \ldots, x_{t-1}\}$$

Let $\mathcal{H}_t$ denote the history that consists of the set of products tested thus far and their performances:

$$\mathcal{H}_t = \{(0,0), (x_1, f(x_1)), \ldots, (x_{t-1}, f(x_{t-1}))\}\,. \tag{3.2}$$

(b) Mass-Production: Upon concluding experimentation, the firm chooses a product to mass-produce. We refer to this phase as period $T+1$. We assume that the firm's profit from mass producing a product x is:

$$\text{Profit from Mass-Producing Product } x = u^{\mathsf{MP}}(f(x))$$

We assume that $u^{\mathsf{MP}}(\cdot)$ is strictly increasing and concave. It is straightforward that the firm's profit from the mass production of a product is increasing in its performance, while concavity implies that the marginal gain in profit is diminishing in the product performance. Further, the concavity of $u^{\mathsf{MP}}(\cdot)$ reflects the higher risk that the firm faces during mass-production as compared to small-batch testing during experimentation.

3.1 Firm's Problem

We assume that the firm chooses a product to experiment, denoted by x_t, by solving the following problem:

$$x_t \in \arg\max_{x \in \mathcal{X}_\infty} \mathbb{E}\left[\max\left\{f(0), f(x_1), f(x_2), \ldots, f(x_{t-1}), f(x)\right\} \middle| \mathcal{H}_t\right] \qquad \text{(PROBLEM EXPN-}t\text{)}$$

That is, the firm chooses a product to maximize the expected ex-post maximum performance in each period t. In period $T+1$, the firm mass-produces a product. The firm's problem during mass-production is as follows:

$$x^{\mathsf{MP}} \in \arg\max_{x \in \mathcal{X}_\infty} \mathbb{E}\left[u^{\mathsf{MP}}\left(f(x)\right) \middle| \mathcal{H}_{T+1}\right]. \qquad \text{(PROBLEM MP)}$$

The firm's objective in PROBLEM EXPN-t corresponds to the best product among all the products that the firm would have tested at the end of period t, reflecting the firm's recourse over the two phases. The firm's objective in PROBLEM MP corresponds to the expected profit from mass-production of product x. As we will soon show, the concavity of $u^{\mathsf{MP}}(\cdot)$ makes mass production of an untested product less appealing.[7]

4 Structural Results

Before we conduct our main analysis, we first present a few structural results below. These results help simplify the firm's decision-making problems shown in PROBLEM EXPN-t and PROBLEM MP above. Specifically, we present the following:

(i) Following any history $\mathcal{H}_t$ in period t, we characterize the firm's posterior belief on $f(x)$, $x \in \mathcal{X}_\infty$, based on the primitives $(m(x), \sigma_0)$ and the history $\mathcal{H}_t$.

(ii) We simplify the firm's problem in its mass-production stage (PROBLEM MP) to show that it suffices to restrict attention to a small set of products.

(iii) We simplify the firm's objective in its experimentation stage at any period t (PROBLEM EXPN-t) as a direct sum of two components: the performance of the best observed product in history thus far and an *expected gain* from the continued experimentation of the next product.

(iv) Under mild regularity conditions on $m(\cdot)$, we provide an upper bound on the expected gain from experimentation in PROBLEM EXPN-t.

The above structural results hold for any general mean function. Further, these results aid us in Sections 5-6, where we develop deeper analytical results for the solutions to PROBLEM EXPN-t and

[7]While it is natural to model the firm's decision-making during its experimentation phase – the product tested in each period of experimentation – as a forward-looking decision-maker, that correctly accounts for how the product tested in each period affects experimentation in future periods, such an analysis is intractable and has been demonstrated to be NP-hard (Garnett, 2023; Liu et al., 2023).

PROBLEM MP for the special case of mean-equivalent product-performance, in which the firm has no reliable prior information about all but the status-quo product.

4.1 Firm's Posterior Belief on $f(x)$ Following any History $\mathcal{H}_t$

From (3.2), recall the set of products tested thus far denoted by $\mathcal{X}_t$ and the firm's history denoted by $\mathcal{H}_t$. For notational convenience, we sort the products in the history $\mathcal{X}_t$ in an increasing sequence. Let $x_t^{(i)} \in \mathcal{X}_t$ denote the product with i^{th} smallest index in period t, i.e.,

$$0 = x_t^{(1)} \leq x_t^{(2)} \leq \ldots x_t^{(t)} \leq 1.$$

The set of products tested $\mathcal{X}_t$ create a partition of the set of entire products $\mathcal{X}_\infty$ with at most t segments. We let $\mathcal{S}_t^{(i)}$ denote the i^{th} segment in period t, i.e.,

$$\mathcal{S}_t^{(i)} = \left[x_t^{(i)}, x_t^{(i+1)}\right] \text{ for } i \in [t-1], \text{ and } \mathcal{S}_t^{(t)} = \left[x_t^{(t)}, 1\right]$$

We will refer to $\mathcal{S}_t^{(t)}$ as the right flank. For any $i \in [t-1]$, define $\epsilon_t^{(i)}$ as follows:

$$\epsilon_t^{(i)} = f\left(x_t^{(i)}\right) - m\left(x_t^{(i)}\right).$$

$\epsilon_t^{(i)}$ is the extent to which the performance of product $x_t^{(i)}$ deviates from its prior mean. Following standard results for a Gaussian process (Williams and Rasmussen, 2006), it can be shown that, for any $x \in \mathcal{X}_\infty$, the firm's posterior belief is as follows: $f(x)\big|\mathcal{H}_t \sim \mathcal{N}\left(\mu_t(x), \sigma_t^2(x)\right)$, where

$$\begin{aligned}
\mu_t(x) &= \begin{cases} m(x) + \epsilon_t^{(i)} + \dfrac{x - x_t^{(i)}}{x_t^{(i+1)} - x_t^{(i)}}\left(\epsilon_t^{(i+1)} - \epsilon_t^{(i)}\right) & x \in \mathcal{S}_t^{(i)},\ i \in [t-1] \\ m(x) + \epsilon_t^{(t)} & x \in \mathcal{S}_t^{(t)} \end{cases}, \text{ and} \\
\sigma_t^2(x) &= \begin{cases} \sigma_0^2 \dfrac{\left(x_t^{(i+1)} - x\right)\left(x - x_t^{(i)}\right)}{x_t^{(i+1)} - x_t^{(i)}} & x \in \mathcal{S}_t^{(i)},\ i \in [t-1] \\ \sigma_0^2\left(x - x_t^{(t)}\right) & x \in \mathcal{S}_t^{(t)} \end{cases}
\end{aligned} \tag{4.1}$$

We provide an illustrative example of the firm's belief in (4.1) and how it evolves with experimentation in Figure 3 below. In this example, we have $m(x) = \sin(2\pi x)$ and $\sigma_0 = 1$. We choose $x_1 = 0.2$, $x_2 = 0.5$ and $x_3 = 0.7$. The orange dotted curve in part (a) is the prior mean ($m(x) = \sin(2\pi x)$), while the red curves in (b), (c), (d) correspond to $\mu_1(x)$, $\mu_2(x)$ and $\mu_3(x)$, respectively; see (4.1). The shaded portion in gray in (a) corresponds to the region $[m(x) + z_{\alpha/2}\sigma_0\sqrt{x}, m(x) + z_{1-\alpha/2}\sigma_0\sqrt{x}]$, while the shaded portions in (b), (c), and (d) correspond to the region $[\mu_t(x) + z_{\alpha/2}\sigma_t(x), \mu_t(x) + z_{1-\alpha/2}\sigma_t(x)]$, with $\alpha = 0.05$ ($-z_{0.025} = z_{0.975} \approx 1.96$). Intuitively, the posterior mean $\mu_t(x)$ in segment $\mathcal{S}_t^{(i)}$ is the sum of the prior mean and the line segment connecting the points $(x_t^{(i)}, f(x_t^{(i)}))$ and $(x_t^{(i+1)}, f(x_t^{(i+1)}))$. The posterior variance $\sigma_t^2(x)$ in segment $\mathcal{S}_t^{(i)}$ is an inverted U-shaped curve with a maximum at the midpoint of the segment. A critical tradeoff for determining the next product to experiment is to balance the "known" and the "unknown": The mean values (i.e., the red curve in Figure 3) represent the expected performance of the untested products derived from the *known* performances of the tested products. The magnitude of the variance (reflected by the shaded area in Figure 3)) represents the level of uncertainty (i.e., the unknown) in the posterior belief; all else equal, products with a higher uncertainty is more attractive for experimentation due to a higher upside for performance.

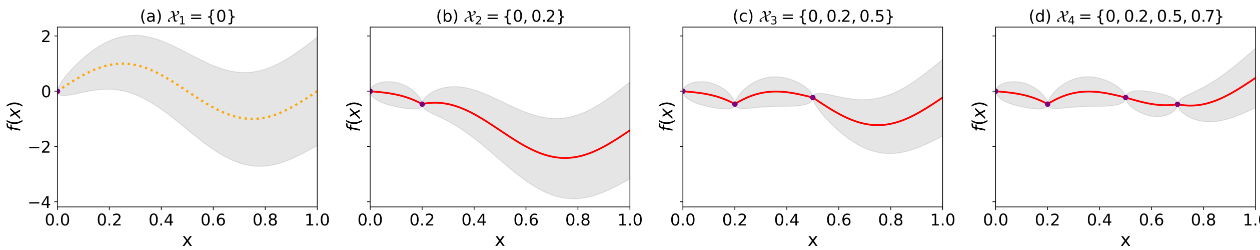

Figure 3. An Illustrative Example on the Firm's Belief Evolution with Experimentation

4.2 Simplifying Problem MP

Recall Problem MP, the firm's problem during mass-production in period $T+1$. The following result shows the firm's optimal decision for Problem MP.

Lemma 1. *Suppose $m(x)$ is convex in x. The product that is mass-produced (x^{MP}) satisfies the following:*

$$x^{\mathsf{MP}} \in \mathcal{X}_{T+1} \cup \mathcal{S}_{T+1}^{(T+1)}.$$

Recall that $\mathcal{X}_{T+1} = \{0, x_1, x_2, \ldots, x_T\}$ denotes the set of all tested products, and $\mathcal{S}_{T+1}^{(T+1)} = \left[x_{T+1}^{(T+1)}, 1\right]$ denotes the right flank. Lemma 1 states that during mass-production, it suffices to restrict attention to the products experimented thus far and the right flank. In other words, any untested product to the left of $x_{T+1}^{(T+1)}$ is not mass-produced. To intuit, consider any untested product x to the left of $x_{T+1}^{(T+1)}$. Say x in the strict interior of segment $\mathcal{S}_{T+1}^{(i)}$ for some $i \leq T+1$. At this value of x, the posterior mean is dominated by one of its boundaries and its variance is higher.[8] The concavity of $u^{\mathsf{MP}}(\cdot)$ penalizes the uncertainty from mass-production of untested products.

An important consequence of Lemma 1 is the following result.

Corollary 1. *Suppose $1 \in \mathcal{X}_{T+1}$. Then, the product that is mass-produced is the product with the best performance among all tested products. Formally,*

$$x^{\mathsf{MP}} \in \arg\max_{x \in \mathcal{X}_{T+1}} f(x).$$

In words, if the optimal experimentation strategy involves testing the (right) boundary product, then

[8]Formally, for $x \in \text{int}(\mathcal{S}_{T+1}^{(i)})$, we have $\mu_{T+1}(x) \leq \max\left\{f(x_{T+1}^{(i)}), f(x_{T+1}^{(i+1)})\right\}$, and $\sigma_{T+1}^2(x) > 0$.

the product that is mass-produced is among the tested products. Stated differently, an untested product is never mass-produced,[9] which resonates with strategies adopted by most practitioners.

4.3 Firm's Objective in Problem Expn-t

Consider any period $t \in [T]$ during the experimentation phase. Recall the firm's problem in Problem Expn-t for any $\mathcal{H}_t$. We can rewrite the firm's problem as follows:

$$\max_{x \in \mathcal{X}_\infty} \left[\underbrace{\max_{i \in [t]} \left\{ f\left(x_t^{(i)}\right) \right\}}_{\substack{\text{performance of the best} \\ \text{product in } \mathcal{X}_t, \triangleq f_t^\star}} + \underbrace{\mathbb{E}\left[\left(f(x) - \max_{i \in [t]} \left\{ f\left(x_t^{(i)}\right) \right\} \right)^+ \big| \mathcal{H}_t \right]}_{\substack{\text{expected gain from continued} \\ \text{experimentation in period } t, \\ \triangleq g_t(x)}} \right] \quad \text{(Problem Expn-}t\text{)} \tag{4.2}$$

We let $f_t^\star$ denote the incumbent – the performance of the best product experimented thus far – in period t, and $g_t(x)$ denote the expected gain from continued experimentation:[10]

$$f_t^\star = \max_{i \in [t]} \left\{ f\left(x_t^{(i)}\right) \right\}, \text{ and } g_t(x) = \mathbb{E}\left[\left(f(x) - f_t^\star\right)^+ \big| \mathcal{H}_t \right]$$

The incumbent $f_t^\star$ is independent of x. Consequently, it suffices to focus on the gain from continued experimentation $g_t(x)$. The firm's problem in period t simplifies to maximizing $g_t(x)$. In our analysis in Sections 5, we will restrict attention to $g_t(x)$, and refer to the problem shown below in (4.3) as Problem Expn-t.

$$\max_{x \in \mathcal{X}_\infty} g_t(x) \quad \text{(Problem Expn-}t\text{)} \tag{4.3}$$

Problem Expn-t in (4.3) is also referred to as the "expected improvement" problem in Bayesian optimization. Much of the literature in this stream has adopted numerical approaches in their search for an optimal solution. As we will soon demonstrate in Section 5, we provide a complete characterization of the optimal solution for this problem for the case of $m(x) = 0$. Futher, in Section 7, we extend it to the case when $m(x)$ is linear.

Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the p.d.f. and c.d.f. of the standard normal distribution, respectively. For any $z \in \Re$, let $H(z)$ denote the following:

$$H(z) = \phi(z) - z\left(1 - \Phi(z)\right). \tag{4.4}$$

In Lemma C.1, we show that $H(\cdot)$ is positive, strictly decreasing, strictly convex, and $\lim_{z \to \infty} H(z) = 0$.

[9] If $1 \in \mathcal{X}_{T+1}$, the right flank is of zero measure, i.e., for all $t \geq 2$, $x_t^{(t)} = 1$, and hence $\mathcal{S}_{T+1}^{(T+1)} = \{1\} \subset \mathcal{X}_{T+1}$.
[10] Indeed, $f_t^\star$ and $g_t(x)$ depend on $\mathcal{H}_t$. We suppress the additional notation as long as no confusion arises.

We can simplify $g_t(x)$ to the following:

$$g_t(x) = \sigma_t(x) H\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right) \tag{4.5}$$

Lemma 2 (Useful Properties of $g_t(x)$)**.** *Fix x. $g_t(x)$ satisfies the following:*

(a) increasing and convex in $\mu_t(x)$.

(b) increasing and convex in $\sigma_t(x)$.

(c) decreasing and convex in $f_t^\star$.

Lemma 2 states that the firm's expected gain from experimenting with product x is (a) higher if the product is expected to have a high performance (effect of $\mu_t(x)$), (b) is not in the vicinity of the tested products thus far (effect of $\sigma_t(x)$), and (c) the performance of the best product so far is low (effect of $f_t^\star$).

Despite Lemma 2, an analytical characterization of an optimal solution to solving PROBLEM EXPN-t for any $m(x)$ remains challenging because $g_t(x)$ is non-convex and is defined piecewise. Below, we present an upper bound for $g_t(x)$ that highlights its structural dependence on the prior mean $m(x)$. This upper bound is tight and presents an intuitive interpretation for the value from continued experimentation.

4.4 An Upper Bound For the Expected Gain in PROBLEM EXPN-t

We first present the notion of non-K-increasing and weak (K_1, K_2)-convexity, introduced by Semple (2007), which generalizes monotonicity and convexity to allow bounded deviations from linearity.

Definition 2. *(a) Fix $K \geq 0$. A continuous function $f(x)$ is non-K-increasing on $[0, U]$ if, for all $x, y \in [0, U]$ with $x \leq y$, it holds that*

$$f(x) + K \geq f(y).$$

(b) Fix $K_1, K_2 \geq 0$. A continuous function $f(x)$ is weakly (K_1, K_2)-convex on the interval $[0, U]$ if and only if for any two points $x, y \in [0, U]$ with $x < y$ and any $\lambda \in [0, 1]$

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)(f(x) + K_1) + \lambda(f(y) + K_2) \tag{4.6}$$

Intuitively, weak (K_1, K_2)-convexity permits a continuous function to deviate slightly above its linear interpolation, with deviations bounded by K_1 and K_2 on the left and right endpoints, respectively. In our context, this concept provides a convenient way to characterize the shape restrictions on the prior mean function $m(x)$ that prevent the posterior mean from overshooting the best observed performance.

Lemma 3. *Consider period $t \in [T]$. For any segment $S_t^{(i)}$ for $i \in [t-1]$ (i.e., any segment except the right flank), define the following:*

$$K_t^{(i)} = f_t^\star - f\left(x_t^{(i)}\right) \text{ and } K_t^{(i+1)} = f_t^\star - f\left(x_t^{(i+1)}\right).$$

For segment $S_t^{(t)}$ (i.e., the right flank), define the following:

$$K_t^{(t)} = f_t^\star - f\left(x_t^{(t)}\right)$$

If $m(x)$ is weakly $\left(K_t^{(i)}, K_t^{(i+1)}\right)$ convex in the segment $\mathcal{S}_t^{(i)}$ for $i \in [t-1]$, and non-$K_t^{(t)}$-increasing in $\mathcal{S}_t^{(t)}$, then

$$g_t(x) \le \frac{\sigma_t(x)}{\sqrt{2\pi}}. \tag{4.7}$$

Equation (4.7) in Lemma 3 provides an upper-bound on the expected gain from continued experimentation: for any candidate $x \in \mathcal{X}_\infty$, the expected gain is bounded from above by $\frac{\sigma_t(x)}{\sqrt{2\pi}}$. Second, this bound is tight in that it is attained when the posterior mean at x coincides with the performance of the best product tested so far (i.e., $m(x) = f_t^\star$). In this case, the expected gain from experimentation is equal to $\frac{\sigma_t(x)}{\sqrt{2\pi}}$. We remark that while Locatelli (1997) derive this upper-bound for the case of $m(x) = 0$, our result in Lemma 3 extends this result to a broader class of mean functions that satisfy weak $\left(K_t^{(i)}, K_t^{(i+1)}\right)$-convexity in all segments but the right flank, and non-$K_t^{(t)}$-increasing in the right flank. In particular, any convex $m(\cdot)$ satisfies the weak convexity requirement.

Crucially, the upper-bound depends only on the posterior uncertainty $\sigma_t(x)$, and not on the performance of any product. From an operational standpoint, the upper-bound provides a useful heuristic for decision-making in such environments: the firm can simply choose products using $\sigma_t(\cdot)$ alone (à la bisection search), without explicitly computing $g_t(x)$ exactly. Below, we provide an example illustrating such an upper bound.

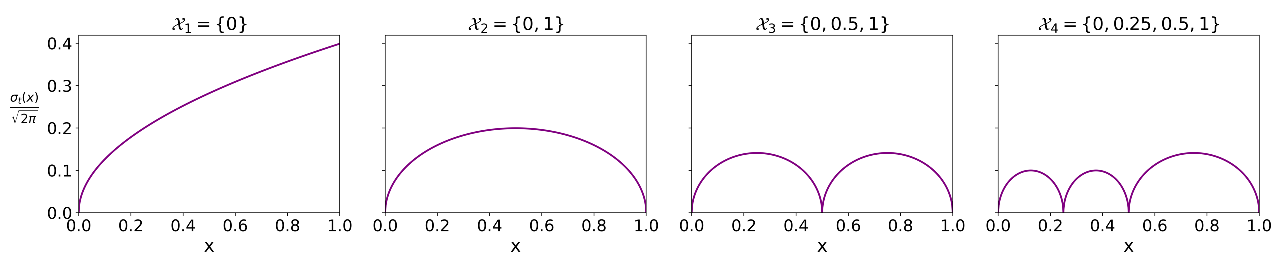

Figure 4. Upper Bound of Expected Gain with $\sigma_0 = 1$, $m(x) = 2x^2$

To develop deeper analytical results on the firm's optimal experimentation strategy, in Sections 5-6, we consider the case of "mean-equivalent" product performance, where the firm has no reliable information apriori about the products (except the status-quo product) and hence is preference-agnostic. As a result, the value of all other products are no different from the status-quo on average; however, they differ in the amount of uncertainty. Without loss of generality, we normalize the mean to zero in this case, and formally state the following assumption.

Assumption 1 (Mean-Equivalent Product Performance)**.** $m(x) = 0$ *for* $x \in \mathcal{X}_\infty$.

Under Assumption 1, $f(x)$, $x \in \mathcal{X}_\infty$, is referred to as a Brownian motion (with zero drift). In period t, using (4.1), $\mu_t(x)$ can be simplified to:

$$\mu_t(x) = \begin{cases} f\left(x_t^{(i)}\right) + \left(\dfrac{f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right)}{x_t^{(i+1)} - x_t^{(i)}}\right)\left(x - x_t^{(i)}\right) & x \in \mathcal{S}_t^{(i)},\ i \in [t-1] \\ f\left(x_t^{(t)}\right) & x \in \mathcal{S}_t^{(t)} \end{cases}, \tag{4.8}$$

The posterior mean in any segment $\mathcal{S}_t^{(i)}$, $i \leq [t-1]$ is the line segment connecting $\left(x_t^{(i)}, f\left(x_t^{(i)}\right)\right)$ and $\left(x_t^{(i+1)}, f\left(x_t^{(i+1)}\right)\right)$. In the right flank, the posterior mean is $f\left(x_t^{(t)}\right)$. The expression for $\sigma_t(x)$ is identical to that in (4.1). Figure 5 provides an illustration of this setting.

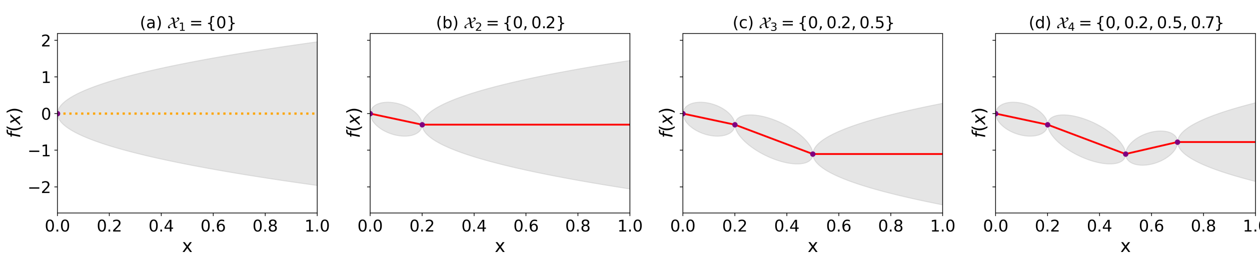

Figure 5. An Illustration of the Firm's Posterior Belief for the Case of $m(x) = 0$.

While Sections 5-6 consider the mean-equivalent setting which substantially simplifies the presentation, our analysis techniques apply for general mean functions. In Section 7, we discuss how to extend our results to settings beyond the mean-equivalent setting.

5 Analysis: Optimal Experimentation and Mass-Production

In this section, we provide a complete characterization of the optimal solutions to Problem Expn-t and Problem MP for the case of mean-equivalent product performance (Assumption 1). We remark that much of the work in this stream has employed numerical approaches in their search for an optimal solution. To our knowledge, we are the first to provide an analytical approach to solving Problem Expn-t and Problem MP, which makes for an important contribution to this stream of work.

5.1 Analysis: Problem Expn-t

Recall the firm's period-t experimentation problem shown in (4.3). We first present the result for period-1 below.

Lemma 4 (Optimal Solution to the Period-1 Problem)**.** *In period-1, the firm's optimal decision is* $x_1^\star = 1$.

Lemma 4 shows that in period-1, the firm's optimal product to experiment is the right extreme. To see this, observe that all products have the same mean, and the standard deviation is increasing in x. From Lemma 2, since the expected gain is increasing in the standard deviation (all else equal), it follows that the optimal decision is $x_1 = 1$.[11]

Next, consider any period $t \geq 2$. From Lemma 4, it follows that for any segment $\mathcal{S}_t^{(i)}$, $i \in [t]$, the boundary points are known. To solve PROBLEM EXPN-t, we proceed in two steps.

(i) Consider a segment $\mathcal{S}_t^{(i)}$ for any $i \in [t]$. We analyze the PROBLEM EXPN-t within this segment, i.e., we solve the following problem:

$$\max_{x \in \mathcal{S}_t^{(i)}} g_t(x). \quad (\text{PROBLEM EXPN-}t\text{-SEG-}(i)) \tag{5.1}$$

We refer to this problem as PROBLEM EXPN-t-SEG-(i). We show that $g_t(x)$, $x \in \mathcal{S}_t^{(i)}$ is unimodal in x. Hence, f.o.c's are necessary and sufficient to identify the optimal solution in this segment.

(ii) To identify the optimal solution to PROBLEM EXPN-t, it suffices to compare the optimal solutions within each segment.

Consider PROBLEM EXPN-t-SEG-(i) in (5.1). Let $x_t^{\star(i)}$ denote a maximizer. Since $g_t(x)$ is continuous in x in this segment, one of the following holds: either f.o.c.'s hold at optimality if the maximizer is in the strict interior of segment $\mathcal{S}_t^{(i)}$, or the maximizer is at the boundary. The derivative of $g_t(x)$ w.r.t. x in this segment can be written as:

$$\frac{dg_t}{dx} = \underbrace{\frac{\partial g_t}{\partial \mu_t}}_{\substack{\text{positive, from} \\ \text{Lemma 2(a)}}} \underbrace{\mu_t'(x)}_{\text{constant, from (4.8)}} + \underbrace{\frac{\partial g_t}{\partial \sigma_t}}_{\substack{\text{positive, from} \\ \text{Lemma 2(b)}}} \underbrace{\sigma_t'(x)}_{\substack{\text{inverted-U shaped} \\ \text{from (4.1)}}} . \tag{5.2}$$

We make a few observations regarding the r.h.s. of (5.2).

(a) Experimentation is Optimal: First, $x_t^{\star(i)}$ is in the strict interior of segment $\mathcal{S}_t^{(i)}$. To see this, observe that at $x = x_t^{(i)}$ (resp., $x = x_t^{(i+1)}$), we have $\sigma_t'(x) = \infty$ (resp., $\sigma_t'(x) = -\infty$). Consequently, the optimal solution in any segment $\mathcal{S}_t^{(i)}$ of non-zero measure strictly involves experimentation with a new product.

(b) Marginal Rate of Substitution and Firm's Tradeoff: Second, since $x_t^{\star(i)}$ is in the strict interior of segment $\mathcal{S}_t^{(i)}$, f.o.c.'s must hold at optimality. From the expression of $\frac{dg_t}{dx}$ from (5.2), at optimality,

[11] Analytically, observe from (4.5) that $g_1(x) = \frac{\sigma_0\sqrt{x}}{\sqrt{2\pi}}$. Hence, $\arg\max_{x \in \mathcal{X}_\infty} g_1(x) = 1$.

we have:

$$\underbrace{\frac{\left(\frac{\partial g_t}{\partial \mu_t}\right)}{\left(\frac{\partial g_t}{\partial \sigma_t}\right)}}_{\text{marginal rate of substitution}} = -\frac{\sigma_t'(x)}{\mu_t'(x)}. \tag{5.3}$$

In spirit, the l.h.s. is akin to the "marginal rate of substitution" (m.r.s.) in an optimal consumption problem: the firm effectively trades-off higher performance (higher mean) for a greater upside from experimentation (higher uncertainty). This tradeoff is central to the firm's experimentation problem. At optimality, the firm optimally balances this trade-off, where the m.r.s. is equal to the quantity in the r.h.s.

Further, the f.o.c. in (5.3) simplifies to the following:

$$\left(f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right)\right)\varphi\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right) = -\sigma_t'(x). \tag{5.4}$$

That is, at any $x_t^{\star(i)}$, (5.4) holds. Here, $\varphi(z) = \frac{1-\Phi(z)}{\phi(z)}$ is the reciprocal of the hazard rate of the standard normal distribution, also known as the Mill's ratio.

(c) Optimal Solution Closer to the Dominant Boundary: Third, recall from (4.8) that $\mu_t'(x) > 0$ in segment $\mathcal{S}_t^{(i)}$ iff $f(x_t^{(i+1)}) > f(x_t^{(i)})$. From (4.1), $\sigma_t(x)$ is increasing (resp., decreasing) in x if $x \le \frac{x_t^{(i)}+x_t^{(i+1)}}{2}$ (resp., $x \ge \frac{x_t^{(i)}+x_t^{(i+1)}}{2}$). It follows from these two observations that

$$\mu_t'(x) > 0 \Leftrightarrow f\left(x_t^{(i+1)}\right) > f\left(x_t^{(i)}\right) \Leftrightarrow x_t^{\star(i)} > \frac{x_t^{(i)} + x_t^{(i+1)}}{2}.$$

That is, any optimal solution to PROBLEM EXPN-t-SEG-(i) is closer to the dominant boundary.

Finally, let $\Delta_t^{(i)}$ and $w_t^{(i)}$ denote the following:

$$\Delta_t^{(i)} \triangleq f_t^\star - f\left(x_t^{(i)}\right), \text{ and } w_t^{(i)} \triangleq \frac{\Delta_t^{(i)}}{\Delta_t^{(i)} + \Delta_t^{(i+1)}}.$$

Observe that $w_t^{(i)} \ge \frac{1}{2}$ iff $f\left(x_t^{(i+1)}\right) \ge f\left(x_t^{(i)}\right)$. Recall from (4.5) that $g_t(x)$ is the product of two terms:

$$g_t(x) = \underbrace{\sigma_t(x)}_{\substack{\text{concave in } x\text{, maximized at} \\ \text{the midpoint of } \mathcal{S}_t^{(i)}}} \times \underbrace{H\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right)}_{\substack{\text{unimodal in } x\text{, maximized at} \\ x = \left(1 - w_t^{(i)}\right)x_t^{(i)} + w_t^{(i)}x_t^{(i+1)}}}.$$

The first term is concave and is maximized at the midpoint of $\mathcal{S}_t^{(i)}$. In Lemma C.5, we show that the second term is unimodal and is maximized at $x = \left(1 - w_t^{(i)}\right)x_t^{(i)} + w_t^{(i)}x_t^{(i+1)}$. It follows that any maximizer $x_t^{\star(i)}$ satisfies:

$$x_t^{\star(i)} \in \mathcal{W}_t^{(i)} \triangleq \left[\min\left\{\frac{x_t^{(i)}+x_t^{(i+1)}}{2}, w_t^{(i)}x_t^{(i)}+\left(1-w_t^{(i)}\right)x_t^{(i+1)}\right\}, \max\left\{\frac{x_t^{(i)}+x_t^{(i+1)}}{2}, w_t^{(i)}x_t^{(i)}+\left(1-w_t^{(i)}\right)x_t^{(i+1)}\right\}\right].$$

Furthermore, in Lemma C.5, we show that $\lim_{x\downarrow x_t^{(i)}} g_t'(x) \geq 0$, while $\lim_{x\uparrow x_t^{(i+1)}} g_t'(x) \leq 0$. Combining all of the above, we have the following result.

Theorem 1 (Unique Maximizer to PROBLEM EXPN-t-SEG-(i))**.** *The f.o.c. in* (5.4) *has exactly one root in the interval* $\mathcal{W}_t^{(i)}$. *Consequently,* $g_t(x)$ *is unimodal in* $x \in \mathcal{S}_t^{(i)}$.

From Theorem 1, it follows that f.o.c's in (5.4) are necessary and sufficient to identify the unique solution to PROBLEM-EXPN-t-SEG-(i).

To solve (ii), i.e., identify the optimal solution to PROBLEM-EXPN-t, it suffices to compare the optimal solutions to PROBLEM-EXPN-t-SEG-(i) for each $i \in [t]$. Consequently, the optimal product to test in period t is:

$$x_t^\star \in \arg\max_{x\in\left\{x_t^{\star(1)}, x_t^{\star(2)}, \ldots, x_t^{\star(t-1)}\right\}} g_t(x).$$

We summarize the optimal experimentation policy for the firm below in Algorithm 1.

Algorithm 1 Optimal Experimentation Policy

Input: $\sigma_0 > 0, T \in \mathbb{N}$
Set: $\mathcal{X}_1 = \{0\}, \mathcal{H}_1 = \{(0,0)\}$
for $t \in [T]$ **do**
  if t = 1 **then**
    $x_t = 1$
  else
    for $i \in [t-1]$ **do**
      $x_t^{\star(i)} = \arg\max_{x\in\mathcal{S}_t^{(i)}} g_t(x)$
    end for
    $i_t^\star \in \arg\max_{i\in[t-1]} g_t\left(x_t^{\star(i)}\right)$
    $x_t^\star = x_t^{\star(i_t^\star)}$, observe $f(x_t)$
    $\mathcal{H}_{t+1} \leftarrow \mathcal{H}_t \cup \{(x_t^\star, f(x_t^\star))\}$
  end if
end for

5.2 Analysis: PROBLEM MP

Upon concluding experimentation, in period $T+1$, the firm chooses a product to mass-produce. Combining Lemma 4 with Corollary 1, it follows that the product that is mass produced is one of the tested products, and is the one associated with the best performance, i.e.,

$$x^{\mathsf{MP}} \in \arg\max_{x\in\mathcal{X}_{T+1}} f(x).$$

5.3 Computational Challenges of the Optimal Experimentation Policy

We now quantify the computational burden of the optimal experimentation policy. Using Lemma 4, the optimal solution in period-1 is $x_1^\star = 1$. At the start of period $t \in \{2, 3, \ldots, T\}$, the set of products searched thus far $\mathcal{X}_t$ create a partition of $\mathcal{X}_\infty$ with $t-1$ disjoint subintervals (segments), i.e., $\left\{\mathcal{S}_t^{(i)}\right\}_{i=1}^{t-1}$. Recall from (4.5) that the expected gain depends on the best observed performance $f_t^\star$. Consider two consecutive periods t and $t+1$. Depending on the outcome in period t, one of the following two cases arises:

(i) $f\left(x_t^\star\right) > f_t^\star \Rightarrow f_{t+1}^\star > f_t^\star$

(ii) $f\left(x_t^\star\right) \leq f_t^\star \Rightarrow f_{t+1}^\star = f_t^\star$

In case (i), the realized performance of the optimal product in period t yields a higher performance exceeding the incumbent. As a result, due to a strict increase in the performance of the incumbent, the expected gain in period $t+1$, $g_{t+1}(x) \neq g_t(x)$ for any $x \in \mathcal{X}_\infty$. In case (ii), the performance of the optimal product in period t fails to yield a higher incumbent. Consequently, $g_{t+1}(x) = g_t(x)$ for x in segments that have not been searched in period t, while $g_{t+1}(x) \neq g_t(x)$ in the two new segments created by $x_t^\star$.

In perod $t \in [T]$, $t \geq 2$ and $f_{t+1}^\star > f_t^\star$, we need to solve $t-1$ optimization problems. That is, we recompute the optimizer of the expected gain $x_t^{\star(i)}$ in segment $\mathcal{S}_t^{(i)}$, for $i \in [t-1]$. The cumulative number of segments where we optimize the expected gain segment-wise over T periods is $\sum_{t=2}^T (t-1) = \frac{T(T-1)}{2}$.

Recall from Theorem 1 that the $x_t^{\star(i)}$ can be numerically identified via bisection search in $\mathcal{W}_t^{(i)} \subset \mathcal{S}_t^{(i)}$. As such, fix a bisection tolerance $\epsilon > 0$, for each period t, $x_t^{\star(i)}$ can be numerically identified via bisection search in $\mathcal{W}_t^{(i)}$. Define $c_t^{(i)}$ as the smallest integer k such that after k bisection steps, the remaining search interval has length at most ϵ, and hence contains an ϵ-accurate maximizer:

$$c_t^{(i)} = \min\left\{k \in \mathbb{N} : 2^{-k}|\mathcal{W}_t^{(i)}| \leq \epsilon\right\} = \left\lceil \log_2\left(\frac{|\mathcal{W}_t^{(i)}|}{\epsilon}\right)\right\rceil,$$

where $|\mathcal{W}_t^{(i)}|$ denotes the length of $\mathcal{W}_t^{(i)}$. Let C_T be the total number of bisection iterations performed over the horizon T. Since period t contains $t-1$ segments, we have

$$C_T = \sum_{t=2}^{T}\sum_{i=1}^{t-1} c_t^{(i)} = \sum_{t=2}^{T}\sum_{i=1}^{t-1}\left\lceil \log_2\left(\frac{|\mathcal{W}_t^{(i)}|}{\epsilon}\right)\right\rceil.$$

Throughout, we impose a non-degeneracy condition on ϵ to ensure that the bisection search does not terminate immediately.

Assumption 2. *There exists a constant $\kappa > 1$ such that*

$$|\mathcal{W}_t^{(i)}| \geq \kappa\epsilon \qquad \textit{for all } t \in \{2, \ldots, T\}, i \in [t-1]$$

Assumption 2 is equivalent to requiring at least one nontrivial bisection step in every segment. Under this condition, the following bound on C_T holds.

Lemma 5. *The cumulative computational complexity of the optimal experimentation policy is quadratic in the horizon. That is:*

$$C_T = \Theta(T^2).$$

Lemma 5 establishes that computing the optimal trajectory becomes increasingly expensive as the horizon grows. Specifically, while the number of segments grows linearly with t, the *average* number of bisection iterations to identify the global optimizer per period also increases linearly. This suggests that as the firm explores more products (larger T), the contemporaneous computation required to identify $x_t^\star$ is more intensive.

6 Segment-Based Search

Motivated by the computational challenges of the optimal experimentation policy in Lemma 5, we propose a heuristic, yet structurally consistent, approach, that we refer to as the *Segment-Based Search*. This method preserves the desirable property of a unique maximizer of the objective function within each segment, while significantly reducing the computational complexity.

6.1 Segment-Based Incumbent, Expected Gain and Firm's Problem

First, we define the *segment-based incumbent* $f_t^{\star(i)}$ as the best observed performances at the two boundaries of the segment $\mathcal{S}_t^{(i)}$:

$$f_t^{\star(i)} = \begin{cases} \max\left\{f\left(x_t^{(i)}\right), f\left(x_t^{(i+1)}\right)\right\} & i \in [t-1] \\ f\left(x_t^{(i)}\right) & i = t \end{cases}$$

For any $x \in \mathcal{S}_t^{(i)}$, we define the *segment-based expected gain* as the expected improvement relative to the associated segment-based incumbent:

$$\tilde{g}_t(x) = \mathbb{E}\left[\left(f(x) - f_t^{\star(i)}\right)^+ |\mathcal{H}_t\right].$$

For any $t \in [T]$, the firm's period-t decision under the segment-based search is to identify $\tilde{x}_t^\star \in \mathcal{X}_\infty$ to maximize the segment-based expected gain as follows

$$\max_{x \in \mathcal{X}_\infty} \tilde{g}_t(x). \quad (\text{PROBLEM SEG-EXPN-}t) \tag{6.1}$$

The main departure from the optimal experimentation policy is in the firm's period-t problem. In contrast to PROBLEM EXPN-t, under segment-based search, the firm solves PROBLEM SEG-EXPN-t. Structurally, $\tilde{g}_t(x)$ parallels the global expected gain $g_t(x)$ defined in (4.5), where the only difference is the replacement of "global" incumbent $f_t^\star$ by the segment-based incumbent $f_t^{\star(i)}$:

$$\tilde{g}_t(x) = \sigma_t(x) H\left(\frac{f_t^{\star(i)} - \mu_t(x)}{\sigma_t(x)}\right). \tag{6.2}$$

6.2 Analysis

In the first period, Problem Seg-Expn-1 is identical to Problem Expn-1. Hence, Lemma 4 holds, i.e., $\tilde{x}_1^\star = x_1^\star = 1$. In period 2, there is only one segment, implying that $\tilde{f}_2^{\star(1)} = f_2^\star$. As a result, Problem Seg-Expn-2 coincides with Problem Expn-2 as well, i.e., $\tilde{x}_2^\star = x_2^\star$. Consider the non trivial period $t \geq 3$. Since the boundary products' performances are observed, Problem Seg-Expn-t can be solved using the same two-step decomposition as in the optimal experimentation policy:

(i) We analyze the subproblem of Problem Seg-Expn-t within the segment $\mathcal{S}_t^{(i)}$, for $i \in [t-1]$, i.e., we solve the following problem:

$$\tilde{x}_t^{\star(i)} \in \arg\max_{x \in \mathcal{S}_t^{(i)}} \tilde{g}_t(x). \quad (\text{Problem Seg-Expn-}t\text{-Seg-}(i)) \tag{6.3}$$

denoted as Problem Seg-Expn-t-Seg-(i). We show that the unimodality of the segment-based expected gain, $\tilde{g}_t(x)$ in $x \in \mathcal{S}_t^{(i)}$, holds. Hence, f.o.c's are necessary and sufficient to identify the optimal solution in this segment.

(ii) To identify the optimal solution to Problem Seg-Expn-t, it suffices to compare the optimal solutions within each segment.

Using f.o.c.'s, any solution to Problem Seg-Expn-t-Seg-(i) must satisfy:

$$\left(f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right)\right)\varphi\left(\frac{f_t^{\star(i)} - \mu_t(x)}{\sigma_t(x)}\right) = -\sigma_t'(x). \tag{6.4}$$

Theorem 2. *For any segment $\mathcal{S}_t^{(i)}$, the f.o.c. in (6.4) has exactly one root in $\mathcal{S}_t^{(i)}$. Consequently, $\tilde{g}_t(x)$ is unimodal in $x \in \mathcal{S}_t^{(i)}$.*

Consequently, the optimal product to test in period t, $\tilde{x}_t^\star$, under the segment-based policy is to compare optimal solutions to Problem Seg-Expn-i for each $i \in [t-1]$:

$$\tilde{x}_t^\star \in \arg\max_{x \in \left\{\tilde{x}_t^{\star(1)}, \tilde{x}_t^{\star(2)}, \ldots, \tilde{x}_t^{\star(t-1)}\right\}} \tilde{g}_t(x).$$

Algorithm 2 summarizes the segment-based search procedure.

Algorithm 2 Segment-Based Search

Input: $\sigma_0 > 0, T \in \mathbb{N}$
Set: $\mathcal{X}_1 = \{0\}, \mathcal{H}_1 = \{(0,0)\}$
for $t \in [T]$ **do**
  if $t = 1$ or 2 **then**
    $\tilde{x}_t^\star = x_t^\star$, $\tilde{i}_t^\star = 1$
  else
    for $i \in \left\{\tilde{i}_t^\star, \tilde{i}_t^\star + 1\right\}$ **do**
      $\tilde{x}_t^{\star(i)} = \arg\max_{x \in \mathcal{S}_t^{(i)}} \tilde{g}_t(x)$
    end for
    $\tilde{x}_t^\star \in \arg\max_{x \in \left\{\tilde{x}_t^{\star(1)}, \tilde{x}_t^{\star(2)}, \ldots, \tilde{x}_t^{\star(t-1)}\right\}} \tilde{g}_t\left(x\right)$
    observe $f(\tilde{x}_t^\star)$
    $\mathcal{H}_{t+1} \leftarrow \mathcal{H}_t \cup \{(\tilde{x}_t^\star, f(\tilde{x}_t^\star))\}$
  end if
end for

6.3 Key Qualitative Difference Between the Two Policies

The optimal experimentation policy evaluates every candidate product against a single threshold, the best performance observed so far. Consider a scenario where the firm discovers an "early winner", i.e., an unusually high outcome. Such a discovery would suppress the expected gain of other candidates, especially those in regions that remain largely unexplored. In this sense, the optimal experimentation policy is prone to over-committing to an early winner and under-investing in experimentation elsewhere.

In contrast, the segment-based search evaluates a product against a segment-based threshold, the best performance observed within a neighborhood (i.e., the segment it lies in). Therefore, a segment with lower boundary values has a higher expected gain, while a segment with high boundary values has a lower expected gain. In this sense, the segment-based search may result in excessive experimentation, including in neighborhoods with lower potential.

6.4 Computational Complexity

We now analyze the computational complexity of this procedure. A critical feature of the segment-based search is the segment based incumbents. Consider two consecutive periods, t and $t+1$. Define the $\left(\tilde{i}_t^\star\right)^{th}$ as the smallest index of the segment that contains the optimal segment-based expected gain in period $t > 2$ as follows

$$\tilde{i}_t^\star \in \arg\max_{j \in [t-1]} \tilde{g}_t\left(\tilde{x}_t^{(\star(j))}\right).$$

In period $t+1$, segment $\mathcal{S}_t^{(\tilde{i}_t^\star)}$ is split into two new segments by $x_t^\star$. Therefore, the product performances at both two boundaries and the segment-based incumbent remain unchanged in the remaining segments. Consequently, $\tilde{g}_t(x) \neq \tilde{g}_{t+1}(x)$ point-wise for all $x \in \mathcal{S}_t^{(\tilde{i}_t^\star)} = \mathcal{S}_{t+1}^{(\tilde{i}_t^\star)} \cup \mathcal{S}_{t+1}^{(\tilde{i}_t^\star+1)}$.

To determine the global optimizer $\tilde{x}_{t+1}^\star$, it is not necessary to reoptimize $\tilde{g}_{t+1}(x)$ segment-wise on $\mathcal{X}_\infty$. Instead, it suffices to: (i) retain the candidate maximizers $\tilde{x}_t^{\star(j)}$ and their associated values for all "unsearched" segments $j \neq \tilde{i}_t^\star$ in period t; (ii) identify the maximizer of $\tilde{g}_{t+1}(x)$ only in $\mathcal{S}_{t+1}^{(\tilde{i}_t^\star)}$ and $\mathcal{S}_{t+1}^{(\tilde{i}_t^\star+1)}$. Thus, PROBLEM SEG-EXPN-t reduces to identifying the segment-wise optimizer of $\tilde{g}_{t+1}(x)$ in two new segments via bisection search and comparing the optimal segment-based expected gain with those in the "unsearched" segments.

To analyze the computational complexity of the segment-based search, analogously to the case of the optimal experimentation policy under Assumption 2, define $\tilde{c}_t^{(i)}$ as the smallest integer k such that after k bisection steps, the remaining search interval has length at most ϵ for segment $\mathcal{S}_t^{(i)}$, and hence contains an ϵ-accurate maximizer:

$$\tilde{c}_t^{(i)} = \min\left\{k \in \mathbb{N} : 2^{-k}\frac{|\mathcal{S}_t^{(i)}|}{2} \leq \epsilon\right\} = \left\lceil \log_2\left(\frac{|\mathcal{S}_t^{(i)}|}{2\epsilon}\right)\right\rceil,$$

The cumulative number of iterations, denoted as $\tilde{C}_T$, is given by:

$$\tilde{C}_T = \sum_{t=2}^{T}\sum_{i=\tilde{i}_{t-1}^\star}^{\tilde{i}_{t-1}^\star+1} \tilde{c}_t^{(i)} = \sum_{t=2}^{T}\sum_{i=\tilde{i}_{t-1}^\star}^{\tilde{i}_{t-1}^\star+1}\left\lceil \log_2\left(\frac{|\mathcal{S}_t^{(i)}|}{2\epsilon}\right)\right\rceil.$$

Lemma 6. *The cumulative computational complexity of the segment-based search is linear in the horizon. That is:*

$$\tilde{C}_T = \Theta(T).$$

Lemma 6 shows that the computational complexity of searching for T periods under the segment based policy is linearly increasing in T. As a result, the average computational complexity is tightly bounded by a constant, implying that as the firm explores more products (larger T), the contemporaneous computation time required to identify an ϵ-accurate optimizer of the segmend based expected gain is invariant to T.

7 Extensions

Thus far, our analysis focuses on a single status-quo product and mean-equivalent product performances as stated in Assumption 1 (i.e., $m(x) \equiv 0$). In this section, we extend the model in two directions. First, we consider the case where the mean product performance is linear, i.e., $m(x) = \alpha x$ for $\alpha \in \Re$. Second, we allow for an arbitrary set of status quo products. Under each extension, we analytically characterize the optimal decision under the optimal experimentation policy and the segment-based search and demonstrate their correspondence to the mean-equivalent setting.

7.1 Extension 1: Linear Mean

Consider the case where $m(x) = \alpha x$ for $\alpha \in \Re$ and any $\sigma_0 > 0$. The induced Gaussian-process prior is a Brownian motion with a drift parameter α and a diffusion parameter σ_0. The status-quo product is identical to that under the mean-equivalent setting, i.e., $\mathcal{H}_1 = \{(0,0)\}$.

In period 1, the expected gain and the segment-based expected gain coincide because the incumbent is identical. By (4.1), the posterior variance is increasing in $x \in \mathcal{X}_\infty$, and the posterior mean is the line-segment connecting $(0,0)$ and $(1,\alpha)$. When $\alpha \geq 0$, both the posterior mean and the posterior variance are increasing in x. Therefore, the expected gain is maximized at the boundary, and hence $x_1^\star = 1$. This conclusion follows directly from Lemma 2 together with Eq. (4.1). When $\alpha < 0$, the posterior mean is decreasing while the posterior variance is increasing in x. Therefore, maximizing the expected gain involves a non-trivial mean-variance trade-off. Lemma B.2 establishes that there exists a unique threshold $\bar{\alpha} < 0$ such that $x_1^\star = 1$ if and only if $\alpha \geq \bar{\alpha}$. If $\alpha < \bar{\alpha}$, Theorem B.2 shows that the maximizer $x_1^\star$ is the unique interior maximizer of $g_1(x)$ and can be computed numerically via bisection search.

For $t \geq 2$, depending on the value of α, we proceed as follows:

- If $\alpha \geq \bar{\alpha}$, the first-period decision is $x_1^\star = 1$. For any subsequent period $t \geq 2$, the set $\mathcal{H}_t$ yields a partition $\mathcal{X}_t$ that splits $\mathcal{X}_\infty$ into $t-1$ segments. In each segment, the two boundary performances are observed. Conditional on these boundary values, both the optimal experimentation policy and the segment-based search reduce to the mean-equivalent case analyzed in Section 5. In this case, for any realization of the true performance mapping under the linear mean, there is a corresponding realized path under the mean-equivalent case, such that the experimentation sequences under the optimal experimentation policy and the segment-based search under those two paths are policy-wise identical.
- If $\alpha < \bar{\alpha}$, for any period $t \in [T]$, there are two possible types of segments: a segment where the performances at both the boundaries are known, or the right flank where the performance at the right boundary is unknown. By Theorem B.2, the expected gain and the segment-based expected gain have a unique maximizer within each segment induced by $\mathcal{H}_t$. The global maximizer is obtained by comparing the performance of these segment-wise maximizers. If there exists $\bar{t} \in [T]$ such that the optimal policy selects the right boundary $x_{\bar{t}}^\star = 1$, then in all remaining periods $t' > \bar{t}$, the search problem reduces to the mean-equivalent case as in Section 5. Otherwise, $x_t^\star$ can be identified by comparing the optimal value of the expected gain and segment-based expected gain among all segments.

7.2 Extension 2: Arbitrary Status-Quo Products

In this subsection, we endow the firm with an arbitrary set of known products under Assumption 1. Formally, let

$$\mathcal{H}_1 = \left\{ \left(x_1^{(1)}, f\left(x_1^{(1)}\right)\right), \ldots, \left(x_1^{(n)}, f\left(x_1^{(n)}\right)\right) \right\},$$

where, without loss of generality, $0 = x_1^{(1)} < \cdots < x_1^{(n)} \leq 1$ and $n \in \mathbb{N}$.[12] The status-quo products induce a partition of set of products into at most n segments, denoted by $\left\{\mathcal{S}_1^{(i)}\right\}_{i=1}^{n}$, and the posterior distribution conditional on $\mathcal{H}_1$ follows from (4.1). We distinguish two cases.

Case 1: $x_1^{(n)} = 1$**.** If the right endpoint is in the status quo, all segments in any period $t \in [T]$ have known boundary performances, i.e., the right flank does not exist. Under Assumption 1, in each period $t \in [T]$, Theorem 1 and 2 imply that both the optimal experimentation policy and the segment-based search have a unique interior maximizer in each segment.

After t additional periods of search, the multiple status-quo case has $n + t$ observed points, which matches the *cardinality* of the history at the beginning of period $n + t$ in the single status-quo setting. However, the two histories are generated by different mechanisms. In the single status-quo setting, all $n + t$ points are sequentially selected by a policy, starting from a single status-quo product. In the multiple status-quo setting, there are n points given exogenously as status-quo, and only the remaining t points are results from sequential search by a policy conditional on those n status quo. Consequently, even though the number of observed points in period t for the multiple status quo setting is identical to that in period $n + t$ for the single status quo case, the resulting sets of locations and realized performances generally differ, and so do the induced segment boundaries. What carries over is structural: the partition contains the same number of segments, each segment has observed performances at both boundaries, and within each segment, the expected gain and the segment-based expected gain have a unique interior maximizer.

Case 2: $x_1^{(n)} < 1$**.** If the right endpoint is not part of the status-quo, the last segment $\mathcal{S}_1^{(n)}$ is a right flank. Under Assumption 1, Lemma 4 and Lemma 2 imply that, on the right flank, both the expected gain and the segment-based expected gain are maximized at the boundary point, i.e., $x_1^{\star(n)} = 1$.

In period t, the partition contains at most $n + t - 1$ segments. If there exists $\bar{t} \in [T]$ such that $x_{\bar{t}}^{\star} = 1$, then from period $\bar{t} + 1$ onward the right endpoint becomes an observed boundary point, and the right flank disappears. The continuation problem for all $t' > \bar{t}$ therefore reduces to the bounded-segment setting of Case 1, with the current set of observed points being a mixture of multiple status quo and the searched trajectory. Otherwise, if $x_t^{\star} \neq 1$ for all $t \in [T]$, then in each period the policy identifies $x_t^{\star}$ by maximizing the expected gain or segment-based expected gain within each segment and then comparing these segment-wise maxima across segments.

Combining the two cases, we obtain the following implications. First, in every period and for every segment induced by the current status quo, both the expected gain and the segment-based expected gain admit a unique maximizer within that segment. Second, once the endpoint 1 is selected (either because it is already in $\mathcal{H}_1$ or because the policy selects it in some $\bar{t}$[13]), from this period onward, the continuation problem has the same structural form as the mean-equivalent single status-quo analysis: the policy compares segment-wise maxima across a collection of segments with known performances at boundaries, and within each segment the maximizer is uniquely characterized.

[12] $\mathbb{N} = \{1, 2, \ldots\}$.

[13] We identify a necessary condition for this scenario, a time dependent threshold on α, in Appendix B, Lemma B.3

8 Numerical Experiments

In this section, we compare the optimal experimentation and the segment-based search policies via numerical experiments. To demonstrate that these two policies are not specifically tailored to a family of unknown true performance mapping, we consider two canonical stochastic processes that correspond to distinct information structures: (i) a Brownian motion, an information environment where the firm only knows a single status quo product whose performance is normalized to zero. This scenario corresponds to our analysis on the main model in Section 5; (ii) a standard Brownian bridge, an information environment where the status quo products are located at the boundaries of the searching space with equal normalized performances, which is a special case of Section 7.2 where $\mathcal{H}_1 = \{(0,0),(1,0)\}$.

8.1 Benchmark

First, we propose a non-adaptive benchmark against which we compare the adaptive policies. Define the location of the (pathwise) maximizer of f as follows:

$$x^\star \in \arg\max_{x\in\mathcal{X}_\infty} f(x).$$

In the first case where f is a Brownian motion, by the "third arcsine law" (Mörters and Peres, 2010), for any $\sigma_0 > 0$, the maximizer $x^\star$ follows the arcsine distribution on $[0,1]$, with density function

$$p_{x^\star}(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

It implies that the distribution of $x^\star$ is invariant to the diffusion parameter. Accordingly, we define a simple probabilistic benchmark that samples products independently from this distribution:

$$x_t \sim p_{x^\star}(\cdot), \quad t \in [T].$$

In the other case, where f is a standard Brownian bridge, for any $\sigma_0 > 0$, the maximizer of a standard Brownian bridge is uniformly distributed (Hobson, 2007). As such, the benchmark for the Brownian bridge, i.e., case (ii), is to independently sample from the Uniform distribution.

8.2 One Normalized Status-Quo

We now compare the performance of the three policies, i.e., benchmark, optimal experimentation, and segment-based search, when f is a Brownian motion (i.e., one normalized status-quo product). Let $\omega \in \Omega$ index a sample path of the Brownian motion, and let $\{\omega_s\}_{s=1}^S$ denote S independent sample paths drawn under the same σ_0. Then, the global maximum of f for path ω is

$$f^\star(\omega) = \max_{x\in\mathcal{X}_\infty} f(x,\omega)$$

For a given policy π and horizon T, let the sequence of evaluation points chosen on path ω be $\{x_t^\pi(\omega)\}_{t=1}^T \subset [0,1]$, and define the best observed performance up to period t on that path by

$$f_t^{\star,\pi}(\omega) = \max_{\tau \in [t]} f\left(x_\tau^\pi(\omega), \omega\right)$$

Then the regret of policy π at period t on path ω is

$$r_t^\pi(\omega) = f^\star(\omega) - f_t^{\star,\pi}(\omega), \quad t \in [T]$$

which measures the optimality gap in the objective value after t evaluations. The simple regret (Bubeck et al., 2009) is computed as the average across the S simulated paths

$$r_t^\pi = \frac{1}{S}\sum_{s=1}^{S} r_t^\pi(\omega_s), \quad t \in [T].$$

A smaller r_t means that the best evaluated point so far achieves a value closer to the pathwise optimum, that is, the policy has identified a candidate whose realized performance is closer (in value) to the true global maximum. In particular, $r_t = 0$ if and only if the policy has already searched at least one maximizer (or an equivalent product attaining the maximum).

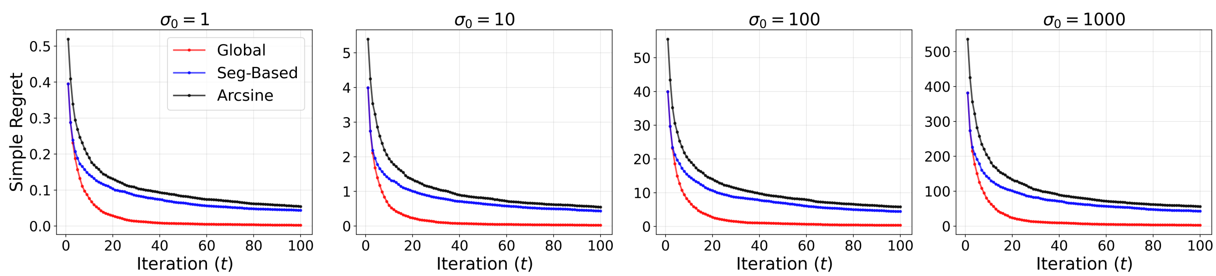

Figure 6. Comparison: Simple Regret with $S = 500, T = 100$

In Figure 6, each dot corresponds to one iteration of a policy; the x-axis reports the iterations, and the y-axis reports the corresponding average (path-wise) simple regret. The benchmark (black) has a higher average simple regret than the other two adaptive policies. The optimal experimentation policy achieves the lowest average simple regret over the horizon. Since the benchmark is nonadaptive, the information from past searches does not affect the firm's future decisions. Moreover, even though the segment-based search, which incorporates only the most "adjacent" information, achieves a low average simple regret, the convergence is significantly slower than the optimal experimentation. This implies that, intuitively, the optimal experimentation policy helps the firm locate the true optimal products more accurately.

Further, while we consider a horizon $T = 100$, in practice, the length of the horizon depends on the ease of experimentation. In apparel, experimentation is likely to last for a much shorter horizon as observing market responses is time-consuming. In such cases, our proposed policies significantly outperform the non-adaptive benchmark.

As search progresses, the set of searched products increases (as more products are added to the history), yielding a finer partition of the search space. Consequently, the optimization problem under the optimal experimentation policy in each period of search takes longer to solve. Since the search cost of the optimal experimentation policy is increasing in the period of search, the total computation time over the entire horizon is higher than that under the segment-based search. We present the average of total computation time in the table below.

σ_0	Average Computation Time (seconds)	
	Optimal Experimentation	Segment-Based
1	6.78	0.42
10	6.65	0.41
100	6.63	0.41
1000	6.55	0.41

To assess the robustness of these search policies, we also report the worst-case simple regret at each period. For policy π in iteration t, the worst-case simple regret is defined as

$$w_t^\pi = \max_{s \in [S]} r_t^\pi(\omega_s)$$

Below, we plot worst-case simple regret over each searching period.

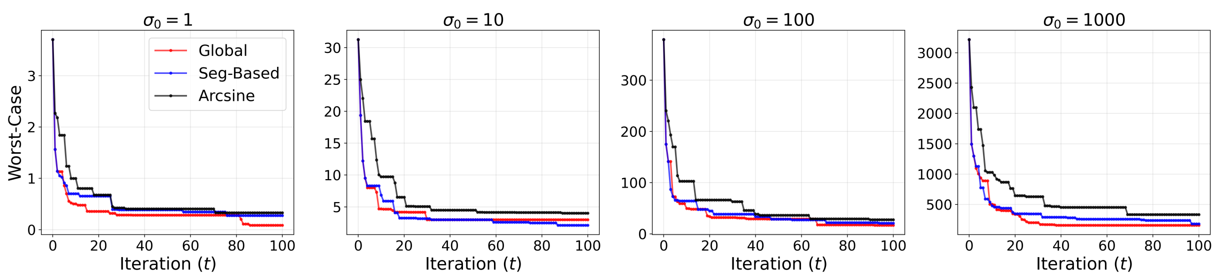

Figure 7. Comparison: Worst Case

In Figure 7, the randomized benchmark exhibits the largest worst-case simple regret throughout most of the horizon, indicating that sampling without using historical information is the least robust. Both adaptive policies substantially reduce the worst-case regret, consistent with learning from historical information.

8.3 Two Status-Quo Products: Boundaries and Equal Performance

When the status-quo has two known products with the same performance at the boundaries, the true function is a realization of a Brownian bridge. In this case, our benchmark is to independently sample from the uniform distribution (Hobson, 2007). The performance and worst-case comparison are presented as follows.

Observe that the performance under simple regret is consistent: the adaptive policies outperform the benchmark. However, in terms of the worst-case, the optimal experimentation policy has a better

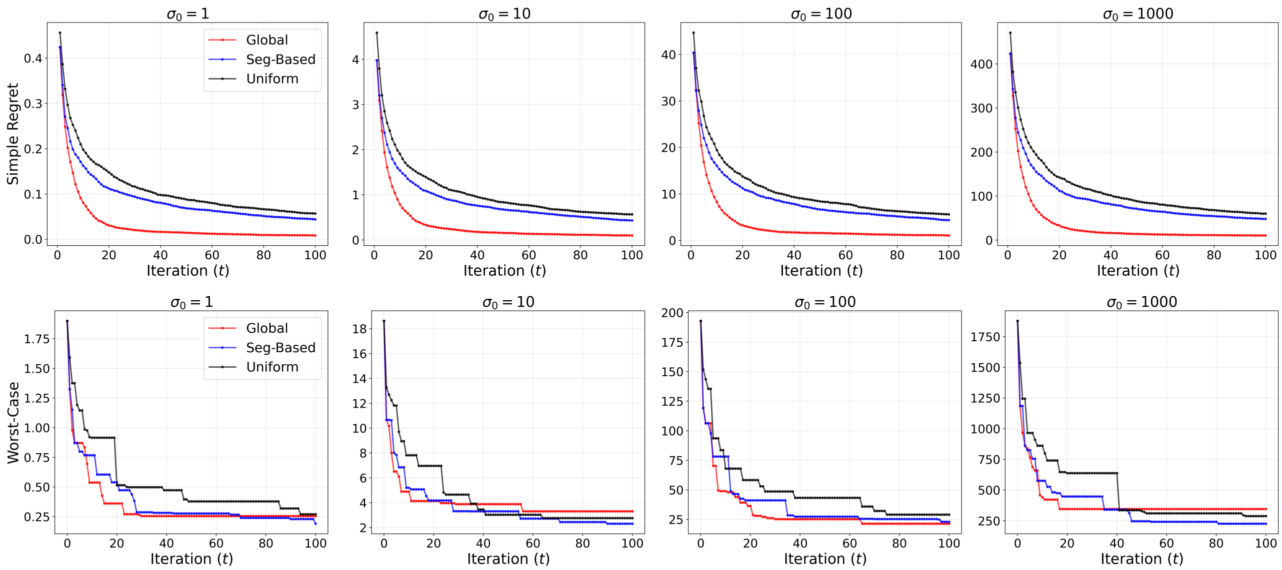

Figure 8. Comparison under Brownian Bridge with $S = 500, T = 100$

worst-case performance than the other two (when $\sigma_0 = 10, 1000$). However, the segment-based policy consistently has better worst-case performance than the benchmark. As such, the segment-based search policy is consistantly more robust than the benchmark policy.

9 Conclusion

Our work in this paper is motivated by practical considerations in new product development. Our model captures several features, e.g., learning spillovers, partial invertability, local unlearnability, etc., that are common in real-world settings, e.g., in apparel, drug development, food and beverage, UI/UX design, etc. From a methodological standpoint, our analysis contributes to the literature in identifying an analytical solution to the optimal experimentation problem. In particular, we show that there is a single product that uniquely offers the greatest potential for improvement. This result is analogous to the index structure in finite-arm models wherein we identify a single optimal bandit for subsequent experimentation. Motivated by the computational burden imposed by the optimal experimentation policy, we propose a structurally consistent, yet computationally easier heuristic that we refer to as segment-based search. We identify the pitfalls of both approaches: the optimal policy is likely to commit to early winners and under-invest in exploration, while the segment-based search is likely to over-invest in exploration, even in regions of low potential. Using numerical experiments, we demonstrate that both approaches significantly outperform a non-adaptive benchmark along two key dimensions: the simple regret, and the worst-case performance. Beyond these, we also extend our model to the case where the prior mean is linear, and the case where the firm may have multiple status-quo products. By providing a novel and attractive framework to study learning spillovers, we view our work as an important building block that future research can build on.

References

Agrawal, Shipra, Vashist Avadhanula, Vineet Goyal, and Assaf Zeevi, "MNL-bandit: A dynamic learning approach to assortment selection," *Operations Research*, 2019, *67* (5), 1453–1485.

Auer, Peter, "Using confidence bounds for exploitation-exploration trade-offs," *Journal of Machine Learning Research*, 2002, *3* (Nov), 397–422.

__ , Nicolo Cesa-Bianchi, and Paul Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, 2002, *47* (2), 235–256.

Baardman, Lennart, Igor Levin, Georgia Perakis, and Divya Singhvi, "Leveraging comparables for new product sales forecasting," *Production and Operations Management*, 2018, *27* (12), 2340–2343.

Baricz, Arpád, "Mills' ratio: Reciprocal convexity and functional inequalities," *arXiv preprint arXiv:1010.3267*, 2010.

Besbes, Omar and Alp Muharremoglu, "On implications of demand censoring in the newsvendor problem," *Management Science*, 2013, *59* (6), 1407–1424.

Bhaskaran, Sreekumar and Sanjiv Erat, "Optimal Prototyping with Noisy Measurements," *Manufacturing & Service Operations Management*, 2025, *27* (1), 322–338.

Broder, Josef and Paat Rusmevichientong, "Dynamic pricing under a general parametric choice model," *Operations Research*, 2012, *60* (4), 965–980.

Bubeck, S et al., "Pure Exploration in Finitely-Armed and Continuous-Armed Bandits problems," in "in" ALT 2009.

Callander, Steven, "Searching and learning by trial and error," *American Economic Review*, 2011, *101* (6), 2277–2308.

Dahan, Ely and Haim Mendelson, "An extreme-value model of concept testing," *Management science*, 2001, *47* (1), 102–116.

den Boer, Arnoud V and N Bora Keskin, "Dynamic pricing with demand learning and reference effects," *Management Science*, 2022, *68* (10), 7112–7130.

Erat, Sanjiv and Stylianos Kavadias, "Sequential testing of product designs: Implications for learning," *Management Science*, 2008, *54* (5), 956–968.

Ferreira, Kris Johnson, David Simchi-Levi, and He Wang, "Online network revenue management using thompson sampling," *Operations research*, 2018, *66* (6), 1586–1602.

Frazier, Peter, Warren Powell, and Savas Dayanik, "The knowledge-gradient policy for correlated normal beliefs," *INFORMS Journal on Computing*, 2009, *21* (4), 599–613.

Garnett, Roman, *Bayesian optimization*, Cambridge University Press, 2023.

Gittins, John, "A dynamic allocation index for the sequential design of experiments," *Progress in statistics*, 1974, pp. 241–266.

Gordon, Robert D, "Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument," *The Annals of Mathematical Statistics*, 1941, *12* (3), 364–366.

Hobson, David, "A short proof of an identity for a Brownian bridge due to Donati-Martin, Matsumoto and Yor," *Statistics & probability letters*, 2007, *77* (2), 148–150.

Huh, Woonghee Tim, Michael Jong Kim, and Meichun Lin, "Uncertain search with knowledge transfer," *Management Science*, 2025.

Kavadias, Stylianos and Karl T Ulrich, "Innovation and new product development: Reflections and insights from the research published in the first 20 years of manufacturing & service operations management," *Manufacturing & Service Operations Management*, 2020, *22* (1), 84–92.

Keskin, N Bora and Assaf Zeevi, "Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies," *Operations research*, 2014, *62* (5), 1142–1167.

Liu, Peng, Haowei Wang, and Wei Qiyu, "Bayesian optimization with switching cost: Regret analysis and lookahead variants," 2023.

Locatelli, Marco, "Bayesian algorithms for one-dimensional global optimization," *Journal of Global Optimization*, 1997, *10* (1), 57–76.

Malladi, Suraj, "An Economic Model of Prior-Free Spatial Search," 2022.

Mitrinovic, Dragoslav S and Petar M Vasic, *Analytic inequalities*, Vol. 1, Springer, 1970.

Mörters, Peter and Yuval Peres, *Brownian motion*, Vol. 30, Cambridge University Press, 2010.

Russo, Daniel, "Simple bayesian algorithms for best arm identification," in "Conference on learning theory" PMLR 2016, pp. 1417–1418.

Ryzhov, Ilya O, "On the convergence rates of expected improvement methods," *Operations Research*, 2016, *64* (6), 1515–1528.

Sampford, Michael R, "Some inequalities on Mill's ratio and related functions," *The Annals of Mathematical Statistics*, 1953, *24* (1), 130–132.

Semple, John, "Note: Generalized notions of concavity with an application to capacity management," *Operations research*, 2007, *55* (2), 284–291.

The Wall Street Journal, "How Shein's On-Demand Model Built a $66 Billion Empire," https://www.wsj.com/video/series/the-economics-of/how-sheins-on-demand-model-built-a-66-billion-empire/81550699-CF45-4E0C-9F9D-DFB42577C929 2023. Accessed: January 30th, 2026.

Thomke, Stefan H, "Managing experimentation in the design of new products," *Management science*, 1998, *44* (6), 743–762.

Thompson, William R, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, 1933, *25* (3/4), 285–294.

Wainwright, Martin J, *High-dimensional statistics: A non-asymptotic viewpoint*, Vol. 48, Cambridge university press, 2019.

Weitzman, Martin L, "Optimal Search for the Best Alternative," *Econometrica*, 1979, *47* (3).

Williams, Christopher KI and Carl Edward Rasmussen, *Gaussian processes for machine learning*, Vol. 2, MIT press Cambridge, MA, 2006.

Yoo, Onesun Steve, Tingliang Huang, and Kenan Arifoğlu, "A theoretical analysis of the lean start-up method," *Marketing Science*, 2021, *40* (3), 395–412.

Zhang, Huanan, Xiuli Chao, and Cong Shi, "Closing the gap: A learning algorithm for lost-sales inventory systems with lead times," *Management Science*, 2020, *66* (5), 1962–1980.

A Proofs of Results in the Main Paper

Proof of Lemma 1. Recall the firm's history $\mathcal{H}_{T+1}$ during mass-production. Since $u^{\mathsf{MP}}(\cdot)$ is concave, from Jensen's inequality, we have:

$$\mathbb{E}\left[u^{\mathsf{MP}}\left(f(x)\right)\middle|\mathcal{H}_{T+1}\right] \leq u^{\mathsf{MP}}\left(\underbrace{\mathbb{E}\left[f(x)\middle|\mathcal{H}_{T+1}\right]}_{\mu_{T+1}(x)}\right).$$

Consider any segment except the right flank, i.e., segment $\mathcal{S}_{T+1}^{(i)}$ for any $i \leq T$. The boundary points of segment $\mathcal{S}_{T+1}^{(i)}$ are $x_{T+1}^{(i)}$ and $x_{T+1}^{(i+1)}$, respectively. From (4.1), for any $x \in \mathcal{S}_{T+1}^{(i)}$:

$$\begin{aligned}
\mu_{T+1}(x) &= m(x) + \epsilon_t^{(i)} + \left(\underbrace{\frac{x - x_T^{(i)}}{x_T^{(i+1)} - x_T^{(i)}}}_{=\lambda}\right)\left(\epsilon_{T+1}^{(i+1)} - \epsilon_{T+1}^{(i)}\right) \\
&= \left(\lambda f\left(x_{T+1}^{(i+1)}\right) + (1-\lambda) f\left(x_{T+1}^{(i)}\right)\right) + \left(\underbrace{m(x) - \left(\lambda m\left(x_{T+1}^{(i+1)}\right) + (1-\lambda) m\left(x_{T+1}^{(i)}\right)\right)}_{\leq 0,\text{ since } m(x) \text{ is convex}}\right) \\
&\leq \max\left\{f\left(x_{T+1}^{(i)}\right), f\left(x_{T+1}^{(i+1)}\right)\right\}.
\end{aligned}$$

Consequently, for $x \leq x_{T+1}^{(T+1)}$, it suffices to restriction to $\mathcal{X}_{T+1}$. Therefore, $x^{\mathsf{MP}} \in \mathcal{X}_{T+1} \cup \mathcal{S}_{T+1}^{(T+1)}$. $\square$

Proof of Lemma 2. Recall from (4.5) that we have

$$g_t(x) = g_t\left(\mu_t(x), \sigma_t(x), f_t^\star\right) = \sigma_t(x) H\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right).$$

(a) To show that $g_t(x)$ is increasing and convex in $\mu_t(x)$, we derive the partial derivatives of $g_t(x)$ w.r.t. $\mu_t(x)$ as follows

$$\begin{aligned}
\frac{\partial g_t(x)}{\partial \mu_t(x)} &= 1 - \Phi\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right) > 0 \\
\frac{\partial^2 g_t(x)}{\partial \mu_t^2(x)} &= \frac{\phi\left(\frac{f_t^\star - \mu_t(x)}{\mu_t(x)}\right)}{\sigma_t(x)} > 0
\end{aligned}$$

(b) To show that $g_t(x)$ is increasing and convex in $\sigma_t(x)$, we derive the partial derivative of $g_t(x)$ w.r.t. $\sigma_t(x)$ as follows

$$\begin{aligned}
\frac{\partial g_t(x)}{\partial \sigma_t(x)} &= H\left(\frac{f_t^\star - \mu_t(x)}{\mu_t(x)}\right) + \frac{f_t^\star - \mu_t(x)}{\mu_t(x)}\left(1 - \Phi\left(\frac{f_t^\star - \mu_t(x)}{\mu_t(x)}\right)\right) \\
&= \phi\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right) \qquad \text{(from (4.4))} \\
\frac{\partial^2 g_t(x)}{\partial \sigma_t^2(x)} &= \frac{\left(f_t^\star - \mu_t(x)\right)^2}{\sigma_t^3(x)} \phi\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right) > 0
\end{aligned}$$

(c) To show that $g_t(x)$ is decreasing and convex in $f_t^\star$, we derive the partial derivative of $g_t(x)$ w.r.t. $f_t^\star$ as follows

$$\frac{\partial g_t(x)}{\partial f_t^\star} = -\left(1 - \Phi\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right)\right) < 0$$

$$\frac{\partial^2 g_t(x)}{\partial (f_t^\star)^2} = \frac{\phi\left(\frac{f_t^\star - \mu_t(x)}{\mu_t(x)}\right)}{\sigma_t(x)} > 0$$

□

Proof of Lemma 3. For any $x \in \mathcal{S}_t^{(i)}, i \in [t-1]$, (i.e., any x not in the right flank), we have

$$\begin{aligned} f_t^\star - \mu_t(x) &= f_t^\star - \left(m(x) + \epsilon_t^{(i)} + \frac{x - x_t^{(i)}}{x_t^{(i+1)} - x_t^{(i)}}\left(\epsilon_t^{(i+1)} - \epsilon_t^{(i)}\right)\right) \\ &= -\left[m(x) - \lambda_t^{(i)}\left(m\left(x_t^{(i+1)}\right) + f_t^\star - f\left(x_t^{(i)}\right)\right) - \lambda_t^{(i+1)}\left(m\left(x_t^{(i+1)}\right) + f_t^\star - f\left(x_t^{(i+1)}\right)\right)\right] \end{aligned}$$

where $\lambda_t^{(i)} = \frac{x_t^{(i+1)} - x}{x_t^{(i+1)} - x_t^{(i)}} = 1 - \lambda_t^{(i+1)}$. Following Definition 2, let

$$K_t^{(i)} = f_t^\star - f\left(x_t^{(i)}\right) \text{ and } K_t^{(i+1)} = f_t^\star - f\left(x_t^{(i+1)}\right).$$

The weak $\left(K_t^{(i)}, K_t^{(i+1)}\right)$ – convexity of $m(x)$ implies that

$$f_t^\star - \mu_t(x) \geq 0 \qquad \text{for } x \in \mathcal{S}_t^{(i)}, i \in [t-1].$$

In Lemma C.1, we show that $H(\cdot)$ is a decreasing function. Consequently, we have:

$$g_t(x) = \sigma_t(x) H\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right) \leq \sigma_t(x) H(0) = \frac{\sigma_t(x)}{\sqrt{2\pi}}. \tag{A.1}$$

For any $x \in \mathcal{S}_t^{(t)}$, (i.e., any x in the right flank), we have:

$$f_t^\star - \mu_t(x) = m\left(x_t^{(t)}\right) + f_t^\star - f\left(x_t^{(t)}\right) - m(x).$$

For $x \in \mathcal{S}_t^{(t)}$, we have $x \geq x_t^{(t)}$. Following Definition 2, let

$$K_t^{(t)} = f_t^\star - f\left(x_t^{(t)}\right).$$

The non-$K_t^{(t)}$increasing of $m(x)$ implies that

$$f_t^\star - \mu_t(x) \geq 0, \qquad \text{for } x \in \mathcal{S}_t^{(t)}.$$

As a result,

$$g_t(x) \leq \frac{\sigma_t(x)}{\sqrt{2\pi}}. \tag{A.2}$$

Combining (A.1) and (A.2), we have the required result. □

Proof of Lemma 4. In the case of $m(x) \equiv 0$, in period 1, we have

$$f_1^\star = 0, \quad \epsilon_1 = f\left(x_1^{(1)}\right) - m\left(x_1^{(1)}\right) = 0 - 0 = 0.$$

It follows that

$$g_1(x) = \sigma_1(x)H(0) = \frac{\sigma_1(x)}{\sqrt{2\pi}} = \frac{\sigma_0}{\sqrt{2\pi}}\sqrt{x}.$$

Consequently, $x_1^\star = 1$. □

Proof of Theorem 1. Depending on the comparison of $f\left(x_t^{(i)}\right)$ and $f\left(x_t^{(i+1)}\right)$, there are two cases to be discussed below:

(a) $f\left(x_t^{(i)}\right) < f\left(x_t^{(i+1)}\right)$,

(b) $f\left(x_t^{(i)}\right) > f\left(x_t^{(i+1)}\right)$.

We apply the following variable transformation as presented in the proof of Lemma C.5:

$$\begin{aligned}
&\Delta_t^{(i)} = f_t^\star - f\left(x_t^{(i)}\right) \geq 0, \quad && \Delta_t^{(i+1)} = f_t^\star - f\left(x_t^{(i+1)}\right) \geq 0,\\
&a_t^{(i)} = x_t^{(i+1)} - x_t^{(i)} > 0, && k_t^{(i)} = \Delta_t^{(i)} - \Delta_t^{(i+1)} = f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right).
\end{aligned}$$

Let

$$s = \frac{x - x_t^{(i)}}{a_t^{(i)}} \in [0,1] \Leftrightarrow x = (1-s)x_t^{(i)} + s x_t^{(i+1)} \in \left[x_t^{(i)}, x_t^{(i+1)}\right].$$

It follows that

$$z_t(s) = \frac{f_t^\star - \mu_t(s)}{\sigma_t(s)} = \frac{\Delta_t^{(i)} - k_t^{(i)} s}{\sigma_t(s)}, \ \sigma_t(s) = \sigma_0\sqrt{a_t^{(i)}}\sqrt{s(1-s)}.$$

Consider case (a). To show that $g_t(s)$ is unimodal in $[0,1]$, from Lemma C.5, it is sufficient to prove that $g_t(s)$ is unimodal in $\left[\frac{1}{2}, \bar{s}_t^{(i)}\right]$, where

$$\bar{s}_t^{(i)} = \frac{\Delta_t^{(i)}}{\Delta_t^{(i)} + \Delta_t^{(i+1)}}.$$

Note that $\forall s \in \left[\frac{1}{2}, \bar{s}_t^{(i)}\right], \quad z_t(s) > 0$. Then, the derivative of $g_t(s)$ w.r.t. s is

$$g_t'(s) = k_t^{(i)}\left(1 - \Phi(z_t(s))\right) + \sigma_t'(s)\phi(z_t(s)) = \phi(z_t(s))\left[k_t^{(i)}\varphi\left(z_t(s)\right) + \sigma_t'(s)\right].$$

Since $g_t(x)$ is continuous, for an interior maximizer, f.o.c's must hold at optimality. Using straightforward algebra

$$g_t'(s) = 0 \Rightarrow k_t^{(i)}\varphi\left(z_t(s)\right) = -\sigma_t'(s).$$

Multiplying a strictly positive term, $z_t(s)$, on both sides, gives us

$$\underbrace{k_t^{(i)} z_t(s)\varphi\left(z_t(s)\right)}_{L(s)} = \underbrace{-z_t(s)\sigma_t'(s)}_{R(s)}. \tag{A.3}$$

To show the unimodality of $g_t(s)$, it suffices to show that (A.3) has a unique solution.

First, we show that $R(s)$ is increasing and $L(s)$ is decreasing in $s \in \left[\frac{1}{2}, \bar{s}_t^{(i)}\right]$. $R(s)$ can be simplified to

$$R(s) = \frac{(2s-1)(\Delta_t^{(i)} - k_t^{(i)} s)}{2s(1-s)}.$$

From straightforward algebra, we have

$$R'(s) = \frac{\left(2\Delta_t^{(i)} - k_t^{(i)}\right) s^2 - 2\Delta_t^{(i)} s + \Delta_t^{(i)}}{2s^2(1-s)^2}.$$

The denominator of $R'(s)$ is positive. The numerator is a quadratic function with the discriminant as follows

$$\left(-2\Delta_t^{(i)}\right)^2 - 4\left(2\Delta_t^{(i)} - k_t^{(i)}\right)\Delta_t^{(i)} = 4\Delta_t^{(i)}\left(k_t^{(i)} - \Delta_t^{(i)}\right) = -4\Delta_t^{(i)}\Delta_t^{(i+1)} \le 0$$

That is, the numerator is a quadratic function of s with a positive leading coefficient and a non-positive discriminant. Therefore, $R'(s) \ge 0$. To show that $L(s)$ is decreasing in s, it suffices to show $L'(s) < 0$. That is,

$$\begin{aligned} L'(s) &= k_t^{(i)} z_t'(s)\left[\varphi\left(z_t(s)\right) + z_t(s)\varphi'\left(z_t(s)\right)\right] \\ &= k_t^{(i)} z_t'(s)\varphi''\left(z_t(s)\right) \quad \text{(using Lemma C.2)} \end{aligned}$$

Since $\varphi(\cdot)$ is convex (from Lemma C.2), we have $L'(s) \le 0$. At $s = \frac{1}{2}$,

$$L\left(\frac{1}{2}\right) > 0 = R\left(\frac{1}{2}\right).$$

At $s = \bar{s}_t^{(i)}$,

$$L\left(\bar{s}_t^{(i)}\right) = k_t^{(i)} z_t(s)\varphi\left(z_t(s)\right) < k_t^{(i)} = R\left(\bar{s}_t^{(i)}\right),$$

where inequality above holds from Lemma C.2. Therefore, (A.3) has a unique solution in $\left[\frac{1}{2}, \bar{s}_t^{(i)}\right]$.

Consider case (b): $f\left(x_t^{(i+1)}\right) < f\left(x_t^{(i)}\right)$. Let $\tilde{k}_t^{(i)} = -k_t^{(i)} > 0$. The f.o.c's can be stated as

$$\tilde{k}_t^{(i)}\varphi\left(z_t(s)\right) = \sigma_t'(s) \tag{A.4}$$

Multiplying $z_t(s)$ on both sides of (A.4) gives us

$$\tilde{k}_t^{(i)} z_t(s)\varphi\left(z_t(s)\right) = z_t(s)\sigma_t'(s) \tag{A.5}$$

To show the unimodality of $g_t(s)$, it suffices to show that (A.5) has a unique interior solution. Following the roadmap of the analysis in case (a), we can symmetrically show that the lhs of (A.5) is increasing in s and the rhs of (A.5) is decreasing in s, for $s \in \left[\bar{s}_t^{(i)}, \frac{1}{2}\right]$.

Combining the results in case (a) and (b), we conclude the proof. □

Proof of Theorem 2. The proof of Theorem 2 is analogous to the proof of Theorem 1 by applying the variable transformation as shown in the proof of Lemma C.6. Then, in the space of $s \in [0, 1]$, we have

$$\tilde{g}_t(s) = \sigma_t(s) H\left(\frac{f_t^{\star(i)} - \mu_t(s)}{\sigma_t(s)}\right).$$

Again, consider the following two cases:

(a) $f\left(x_t^{(i)}\right) < f\left(x_t^{(i+1)}\right)$,

(b) $f\left(x_t^{(i)}\right) > f\left(x_t^{(i+1)}\right)$.

In case (a), we have $f_t^{\star(i)} = f\left(x_t^{(i+1)}\right)$. From Lemma C.6, the maximizer must be in $\left[\frac{1}{2}, 1\right]$. First, we have the f.o.c's of $\tilde{g}_t(s)$ w.r.t. s as follows

$$k_t^{(i)} \varphi(z_t(s)) = -\sigma_t'(s). \tag{A.6}$$

Since

$$\begin{aligned}\lim_{s\uparrow 1} \frac{f_t^{\star(i)} - \mu_t(s)}{\sigma_t(s)} &= \lim_{s\uparrow 1} \frac{\Delta_t^{(i)} - k_t^{(i)} s}{\sigma_0 \sqrt{a_t^{(i)}}\sqrt{s(1-s)}} \\ &= \lim_{x\uparrow 1} \frac{k_t^{(i)}}{\sigma_0\sqrt{a_t^{(i)}}} \sqrt{\frac{1-s}{s}} = 0,\end{aligned}$$

it follows that $\lim_{s\uparrow 1} \tilde{g}_t'(s) = \lim_{s\uparrow 1} k_t^{(i)}\varphi(0) + \lim_{s\uparrow 1}\sigma_t'(s) = -\infty$, and $\tilde{g}_t'\left(\frac{1}{2}\right) = \frac{k_t^{(i)}}{\sigma_0\sqrt{a_t^{(i)}}} > 0$. It is sufficient to show that $\tilde{g}_t(s)$ is unimodal in $\left(\frac{1}{2}, 1\right)$. From Lemma C.6, we have $z_t(s) > 0$ for $s \in \left(\frac{1}{2}, 1\right)$. Multiplying $z_t(s)$ on both sides, we have

$$\underbrace{k_t^{(i)} z_t(s)\varphi\left(z_t(s)\right)}_{L(s)} = \underbrace{-z_t(s)\sigma_t'(s)}_{R(s)}.$$

It follows that

$$R(s) = k_t^{(i)}\left(1 - \frac{1}{2s}\right),$$

which is increasing in s. From Lemma C.2, we have

$$L'(s) = k_t^{(i)} z_t'(s)\varphi(z_t(s)) < 0.$$

Now, we verify the limits at two boundaries:

$$L\left(\frac{1}{2}\right) > 0 = R\left(\frac{1}{2}\right), \quad \lim_{s\uparrow 1} L(s) = -\infty < \frac{k_t^{(i)}}{2} = R\left(\frac{1}{2}\right).$$

Consequently, we conclude that there is a unique root for Eq. (A.6).

In case (b), we have $f_t^{\star(i)} = f\left(x_t^{(i)}\right)$. Let $\tilde{k}_t^{(i)} = -k_t^{(i)} > 0$. Analogously, it suffices to show that the following equation has a unique interior solution:

$$\tilde{k}_t^{(i)} z_t(s)\varphi\left(z_t(s)\right) = z_t(s)\sigma_t'(s). \tag{A.7}$$

Symetrically, we are able to show that the lhs of (A.7) is increasing in s and the rhs of (A.7) is decreasing in s, for $s \in [0, \frac{1}{2}]$.

Combining the results in case (a) and (b), we conclude the proof. □

Proof of Lemma 5. First, we identify a lower bound of C_T. Note that

$$c_t^{(i)} = \left\lceil \log_2\left(\frac{\mathcal{W}_t^{(i)}}{\epsilon}\right)\right\rceil \geq 1.$$

It follows that

$$C_T = \sum_{t\in[T-1]}\sum_{i\in[t]}\left\lceil \log_2\left(\frac{\mathcal{W}_t^{(i)}}{\epsilon}\right)\right\rceil \geq \frac{T(T-1)}{2} = \Omega(T^2).$$

Next, we derive the upper bound of $\sum_{i\in[t]} c_t^{(i)}$, for $t \in [T-1]$.

$$\begin{aligned}
\sum_{i\in[t]} c_t^{(i)} &= \sum_{i\in[t]}\left\lceil \log_2\left(\frac{\mathcal{W}_t^{(i)}}{\epsilon}\right)\right\rceil \\
&\leq \sum_{i\in[t]}\left(\log_2\left(\frac{\mathcal{W}_t^{(i)}}{\epsilon}\right)+1\right) \\
&= t\log_2\frac{1}{\epsilon} + t + \sum_{i\in[t]}\log_2 \mathcal{W}_t^{(i)} \\
&\leq t\log_2\frac{1}{\epsilon} + t + t\log_2\left(\frac{1}{t}\right) \qquad \text{(from Jensen's inequality)} \\
&= t\left(\log_2\frac{1}{\epsilon t}+1\right).
\end{aligned}$$

Then, we have an upper bound of C_T as

$$\begin{aligned}
C_T &= \sum_{t\in[T-1]}\sum_{i\in[t]} c_t^{(i)} \\
&\leq \log_2\frac{1}{\epsilon}\sum_{t\in[T-1]} t - \sum_{t\in[T-1]} t\log_2 t + \frac{T(T-1)}{2}.
\end{aligned}$$

Since $x\log_2 x$ is increasing in $x \in [1, +\infty]$, by Lemma C.4, we have

$$\begin{aligned}
\sum_{t\in[T-1]} t\log_2 t &\geq \int_1^{T-1} t\log_2 t dt \\
&= \frac{1}{\log 2}\left(\frac{(T-1)^2}{2}\log(T-1) - \frac{(T-1)^2}{4} + \frac{1}{4}\right).
\end{aligned}$$

The last equality holds since $\int x\log_2 x dx = \frac{1}{\log 2}\left(\frac{x^2}{2}\log x - \frac{x^2}{4}\right)$, where $\log(\cdot)$ is the natural log.

Thus,

$$
\begin{aligned}
C_T &\leq \left(\log_2 \frac{1}{\epsilon} + 1\right) \frac{T(T-1)}{2} - \frac{(T-1)^2}{2} \log_2(T-1) + \frac{(T-1)^2}{4 \log 2} \\
&= \mathcal{O}(T^2).
\end{aligned}
$$

In conclusion,

$$
\Omega(T^2) \leq C_T \leq \mathcal{O}(T^2).
$$

□

Proof of Lemma 6. Now, we identify a lower and a upper bound for $\tilde{C}_T$.

For the segment-based algorithm, the bisection search is necessary for any period $t \geq 2$. In the second period, the bisection search is only executed in $[0,1]$. Therefore,

$$
\begin{aligned}
\tilde{C}_T &= \sum_{t=2}^{T} \sum_{i=\tilde{i}^\star_{t-1}}^{\tilde{i}^\star_{t-1}+1} \tilde{c}_t^{(i)} \\
&= \tilde{c}_2^{(1)} + \sum_{t=3}^{T} \sum_{i=\tilde{i}^\star_{t-1}}^{\tilde{i}^\star_{t-1}+1} \tilde{c}_t^{(i)} \\
&\geq 2T - 3.
\end{aligned}
$$

Note that,

$$
\tilde{c}_2^{(1)} \leq \log_2\left(\frac{1}{\epsilon}\right) + 1.
$$

And for any $t \geq 3$, the maximum length will be no greater than $1-(t-2)\epsilon$. As such,

$$
c_{\texttt{seg}}(t) \leq 2\left(\log_2 \frac{1-(t-2)\epsilon}{\epsilon} + 1\right).
$$

As a result, we have an upper bound for $\tilde{C}_T$ as follows

$$
\begin{aligned}
\tilde{C}_T &\leq \log_2 \frac{1}{\epsilon} + 2T - 3 + 2\sum_{t=3}^{T} \log_2\left(\frac{1-(t-2)\epsilon}{\epsilon}\right) \\
&= (2T-3)\left(\log_2 \frac{1}{\epsilon} + 1\right) + 2\sum_{t=3}^{T} \log_2\left(1-(t-2)\epsilon\right) \\
&\leq (2T-3)\left(\log_2 \frac{1}{\epsilon} + 1\right) + \frac{2}{\epsilon}\int_0^{(T-2)\epsilon} \log_2(1-s)ds \\
&= (2T-3)\left(\log_2 \frac{1}{\epsilon} + 1\right) + \frac{2}{\epsilon \log 2}\left(-(1-(T-2)\epsilon)\log(1-(T-2)\epsilon) + (2-T)\epsilon\right).
\end{aligned}
$$

The last inequality holds since $\int_0^a \log_2(1-u)du = \frac{1}{\log 2}\left(-(1-a)\log(1-a) - a\right)$. Therefore,

$$
\Omega(T) \leq \tilde{C}_T \leq \mathcal{O}(T).
$$

□

B Supporting Results for Section 7

B.1 Extension 1: Linear Mean

The following lemma establishes the posterior mean for the linear prior mean and single status quo case.

Lemma B.1. *In any period $t \in [T]$ with the associated history $\mathcal{H}_t$, under the linear mean $m(x) = \alpha x$, the posterior mean conditional on $\mathcal{H}_t$ at the beginning of period t is*

$$\mu_t(x) = \begin{cases} f\left(x_t^{(i)}\right) + \dfrac{x - x_t^{(i)}}{x_t^{(i+1)} - x_t^{(i)}} \left(f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right)\right), & x \in \mathcal{S}_t^{(i)}, i \in [t-1] \\ \alpha\left(x - x_t^{(i)}\right) + f\left(x_t^{(i)}\right), & x \in \mathcal{S}_t^{(t)}. \end{cases}$$

Proof of Lemma B.1. The result of this lemma is from the direct application of Eq. (4.1) to the linear prior mean function, i.e., $m(x) = \alpha x$. There are t searched products in $\mathcal{H}_t$, i.e., $\mathcal{X}_t$, which is a partition of $\mathcal{X}_\infty$. When $x_t^{(t)} = 1$ (resp. $x_t^{(t)} < 1$), $\mathcal{X}_\infty$ is split into $t-1$ (resp. t) segments.

Suppose $x_t^{(t)} = 1$. There are $t-1$ segments for which the product performances at both boundaries are known. By Eq. (4.1), for $x \in \mathcal{S}_t^{(i)}$, $i \in [t-1]$, we have

$$\epsilon_t^{(i)} = f\left(x_t^{(i)}\right) - m\left(x_t^{(i)}\right).$$

It follows that

$$\begin{aligned} \mu_t(x) &= m(x) + \epsilon_t^{(i)} + \frac{x - x_t^{(i)}}{x_t^{(i+1)} - x_t^{(i)}} \left(\epsilon_t^{(i+1)} - \epsilon_t^{(i)}\right) \\ &= \alpha x + f\left(x_t^{(i)}\right) - \alpha x_t^{(i)} + \frac{x - x_t^{(i)}}{x_t^{(i+1)} - x_t^{(i)}} \left(f\left(x_t^{(i+1)}\right) - \alpha x_t^{(i+1)} - f\left(x_t^{(i)}\right) + \alpha x_t^{(i)}\right) \\ &= f\left(x_t^{(i)}\right) + \frac{x - x_t^{(i)}}{x_t^{(i+1)} - x_t^{(i)}} \left(f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right)\right). \end{aligned}$$

Suppose $x_t^{(t)} < 1$. There are t segments. Again, segments $\mathcal{S}_t^{(i)}$ for $i \in [t-1]$ are those with product performances known at both boundaries. Therefore, the posterior mean $\mu_t(x)$ for $x \in \mathcal{S}_t^{(i \in [t-1])}$ is identical to the structure of the result above. The remaining segment $\mathcal{S}_t^{(t)}$ is the right flank where the product performance at the right boundary is unknown. For $x \in \mathcal{S}_t^{(t)}$, we have

$$\begin{aligned} \mu_t(x) &= m(x) - \epsilon_t^{(t)} \\ &= \alpha x + f\left(x_t^{(t)}\right) - \alpha x_t^{(t)} \\ &= \alpha\left(x - x_t^{(t)}\right) + f\left(x_t^{(t)}\right). \end{aligned}$$

Combining these two cases, we conclude the proof. □

In the first period, $f_1^\star = 0$ and $\mu_1(x) = \alpha x$, $\sigma_1(x) = \sigma_0\sqrt{x}$. As such, the expected gain function in the first period is

$$g_1(x) = \sigma_1(x)H(z_1(x)) = \sigma_0\sqrt{x}H\left(-\frac{\alpha}{\sigma_0}\sqrt{x}\right).$$

As such, the derivative of $g_1(x)$ w.r.t. x is

$$g_1'(x) = \underbrace{(1-\Phi(z_1(x)))}_{①}\underbrace{\left(\frac{\sigma_1'(x)}{\varphi\left(z_1(x)\right)}+\alpha\right)}_{②},$$

where $z_1(x) = -\frac{\alpha}{\sigma_0}\sqrt{x}$.

For any $\alpha \geq 0$, ① and ② are all non-negative. Therefore, $g_1'(x) \geq 0$. In this case, $x_1^\star = 1$ both for the optimal experimentation policy and the segment-based search.

In the case of $\alpha < 0$, ① ≥ 0, but the sign of ② is indefinite. Consequently, the sign of $g_1'(x)$ is identical to the sign of ②.

Now, we show that ② is decreasing in x. For notation ease, let $\mathfrak{m}(x) = \frac{1}{\varphi(x)}$ be the inverse Mills Ratio. Then, using the following variable transformation

$$s = \sqrt{x},$$

we have

$$\frac{d②}{ds} = \frac{\mathfrak{m}(z_1(s))}{s^2}\left[\left(\mathfrak{m}(z_1(s)) - z_1(s)\right)z_1'(s)s - 1\right].$$

Since $\frac{\mathfrak{m}(z_1(s))}{s^2} > 0$, the sign of $\frac{d②}{ds}$ is identical to the sign of $\left(\mathfrak{m}(z_1(s)) - z_1(s)\right)z_1'(s)s - 1$.

From Lemma C.2, $\mathfrak{m}(z) - z < \frac{1}{z}$ and it is straightforward to verify that $z_1'(s)s = z_1(s) \geq 0$. We have

$$\left(\mathfrak{m}(z_1(s)) - z_1(s)\right)z_1'(s)s - 1 < 0.$$

As a result, ② is decreasing in $s \in [0,1]$. At the left boundary, we have

$$\lim_{s\downarrow 0} ② = +\infty.$$

Consequently, ② crosses 0 from above at most once. Define $G(\alpha) \equiv ②\Big|_{x=1} = \frac{\sigma_0}{2\varphi\left(-\frac{\alpha}{\sigma_0}\right)} + \alpha = \frac{\sigma_0}{2}\mathfrak{m}\left(-\frac{\alpha}{\sigma_0}\right) + \alpha$. Then, ② crosses 0 from above exactly once if and only if $G(\alpha) < 1$.

Lemma B.2. *$G(\alpha)$ has a unique root $\bar{\alpha} < 0$.*

Proof of Lemma B.2. The derivative of $G(\alpha)$ w.r.t. α is

$$G'(\alpha) = -\frac{\mathfrak{m}'\left(-\frac{\alpha}{\sigma_0}\right)}{2} + 1 > 0.$$

The inequality holds due to Lemma C.3.

Further, $G(0) > 0$, and $\lim_{\alpha \to -\infty} G(\alpha) = \lim_{\alpha \to -\infty} \frac{\alpha}{2} < 0$ due to $\mathfrak{m}(z) \sim z$. Therefore, $\bar{\alpha}$ uniquely solves

$$\frac{\sigma_0}{2\varphi\left(-\frac{\bar{\alpha}}{\sigma_0}\right)} + \bar{\alpha} = 0. \tag{B.1}$$

□

Theorem B.1. *For $\alpha \geq \bar{\alpha}$, $x_1^\star = 1$. Otherwise, $x_1^\star = \tilde{x}_1^\star$ is the unique root that solves*

$$\frac{\sigma_1'(x)}{\varphi\left(z_1(x)\right)} + \alpha = 0.$$

Proof of Theorem B.1. This theorem is directly implied by Lemma B.2. In the first period, there is only one segment. Therefore, the expected gain and the segment-based expected gain are identical, i.e., $x_1^\star = \tilde{x}_1^\star$. Suppose $\alpha \geq \bar{\alpha}$, $g_1(x)$ is increasing in x, implying that $x_1^\star = 1$. Otherwise,② has a unique root in $\mathcal{X}_\infty$, which solves

$$② = \frac{\sigma_1'(x)}{\varphi\left(z_1(x)\right)} + \alpha = 0.$$

□

B.2 Extension 2: Arbitrary Status-Quo Products

Consider the linear mean with a status quo consisting of n products, $\mathcal{H}_1 = \left\{\left(x_1^{(1)}, f\left(x_1^{(1)}\right)\right), \dots, \left(x_1^{(n)}, f\left(x_1^{(n)}\right)\right)\right\}$ where $0 = x_1^{(1)} < x_1^{(2)} < \cdots < x_1^{(n)} \leq 1$. Now, we show that the expected gain and the segment-based expected gain have a unique maximizer in each segment, under the linear prior mean and any status quo.

Theorem B.2. *For $t \in [T]$, $g_t(x)$ and $\tilde{g}_t(x)$ has a unique maximizer in segment $\mathcal{S}_t^{(i)}$, $i \in [n+t-1]$.*

Proof of Theorem B.2. For all segments with known boundary performances, the expected gain under the global search and segment-based search shares the same structure as those in the mean-equivalent case. To prove this theorem, it suffices to show that there is a unique optimizer for the expected gain on the right flank, i.e., $\mathcal{S}_t^{(n+t-1)}$, when $\alpha < 0$ for any $t \in [T]$ [14].

Consider $x \in \mathcal{S}_t^{(t+n-1)} = \left[x_t^{(t+n-1)}, 1\right]$, we have the expected gain under global search as

$$g_t(x) = \sigma_t(x) H\left(\underbrace{\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}}_{z_t(x)}\right),$$

where

$$\mu_t(x) = \alpha\left(x - x_t^{(t+n-1)}\right) + f\left(x_t^{(t+n-1)}\right)$$

$$\sigma_t(x) = \sigma_0\sqrt{x - x_t^{(t+n-1)}}$$

[14]The case of $\alpha \geq 0$ is trivial: $x_t^{\star(n+t-1)} = \tilde{x}_t^{\star(n+t-1)} = 1$

and

$$\begin{aligned}
g_t'(x) &= \sigma_t'(x)\phi(z_t(x)) + \alpha\,(1-\Phi(z_t(x))) \\
&= (1-\Phi(z_t(x)))\left(\frac{\sigma_t'(x)}{\varphi(z_t(x))}+\alpha\right).
\end{aligned}$$

Note that the first term $(1-\Phi(z_t(x)))$ is positive. Therefore, it is sufficient to prove $\frac{\sigma_t'(x)}{\varphi(z_t(x))}$ cross $-\alpha$ at most once. Explicitly, we have

$$\frac{\sigma_t'(x)}{\varphi(z_t(x))} = \frac{\sigma_0}{2\sqrt{x-x_t^{(t+n-1)}}}\frac{\phi(z_t(x))}{1-\Phi(z_t(x))}.$$

We perform the following variable transformation:

$$s = \sqrt{x-x_t^{(t+n-1)}} \in \left[0, \sqrt{1-x_t^{(t+n-1)}}\right].$$

For notation ease, we let

$$\Delta = f_t^\star - f\left(x_t^{(n+t-1)}\right) \geq 0, \quad a = 1-x_t^{(n+t-1)} > 0.$$

Then,

$$z_t(s) = \frac{k}{s}+\beta s, \quad k = \frac{\Delta}{\sigma_0}, \beta = \frac{-\alpha}{\sigma_0}.$$

Now, we show that $F(s) = \frac{\phi(z_t(s))}{s(1-\Phi(z_t(s)))}$ decreases in $s \in [0,+\infty)$. Let the inverse Mills ratio be $\mathfrak{m}(z) = \frac{1}{\varphi(z)}$. Then, we have

$$\begin{aligned}
F'(s) &= \frac{\mathfrak{m}'(z_t(s))z_t'(s)s - \mathfrak{m}(z_t(s))}{s^2} \\
&= \frac{\mathfrak{m}(z_t(s))\left[\mathfrak{m}(z_t(s)) - z_t(s)\right]z_t'(s)s - \mathfrak{m}(z_t(s))}{s^2} \\
&= \frac{\mathfrak{m}(z_t(s))}{s^2}\left[\left(\mathfrak{m}(z_t(s)) - z_t(s)\right)z_t'(s)s - 1\right].
\end{aligned}$$

Since $z_t'(s)s = \beta s - \frac{k}{s} \leq z_t(s)$, it is trivial that $z_t'(s)s \leq 0$. As a result, $F'(s) < 0$.

In the case of $z_t'(s)s > 0$, we have

$$\left(\mathfrak{m}(z_t(s)) - z_t(s)\right)z_t'(s)s - 1 < \frac{1}{z_t(s)}z_t'(s)s - 1 < \frac{1}{z_t(s)}z_t(s) - 1 = 0.$$

The first inequality is from Lemma C.2. As a result, $F(s)$ is decreasing in $s > 0$. That is, for $x \in \mathcal{S}_t^{(n+t-1)}$, $g_t(x)$ has a unique interior optimizer if and only if $\frac{\sigma_t'(1)}{\varphi(z_t(1))} < -\alpha$. Otherwise, $g_t(x)$ is maximized at $x=1$. As for the segment-based expected gain, $z_t'(s)s = \beta s = z_t(s)$, the same conclusion follows.

□

Lemma B.3. *For $t \in [T]$ and $x \in \mathcal{S}_t^{(n+t-1)}$, there exists a unique $\bar{\alpha}_{t,g}$ (resp. $\bar{\alpha}_{t,s}$) for the optimal experimentation policy (resp. segment-based search) such that $x_t^{\star(n+t-1)} = 1$ (resp. $\tilde{x}_t^{\star(n+t-1)} = 1$) if and only if $\alpha > \bar{\alpha}_{t,g}$ (resp. $\alpha > \bar{\alpha}_{t,s}$).*

Proof of Lemma B.3. To show this Lemma, it is equivalent to show that there exists a unique $\alpha < 0$ solving

$$\frac{\sigma_t'(1)}{\varphi\left(z_t(1)\right)} = -\alpha.$$

For the optimal experimentation policy, since $z_t(1) = \frac{\Delta}{\sigma_0\sqrt{1-x_t^{(n+t-1)}}} - \alpha\frac{\sqrt{1-x_t^{(n+t-1)}}}{\sigma_0}$, and $\sigma_t'(1) = \frac{\sigma_0}{2\sqrt{1-x_t^{(n+t-1)}}}$, for notation ease, let

$$z_0 = \frac{\Delta}{\sigma_0\sqrt{1-x_t^{(n+t-1)}}}, \quad u = -\alpha\frac{\sqrt{1-x_t^{(n+t-1)}}}{\sigma_0}.$$

Then,

$$\frac{\sigma_t'(1)}{\varphi\left(z_t(1)\right)} = -\alpha \Leftrightarrow \mathfrak{m}(z_0+u) - 2u = 0 \tag{B.2}$$

where $\mathfrak{m}(z) = \frac{1}{\varphi(z)}$. It suffices to show that $\mathfrak{m}(z_0+u) - 2u$ crosses 0 from above exactly once. Define $D(u) = \mathfrak{m}(z_0+u) - 2u$. Next, we show that $D(u)$ is decreasing in u. From Lemma C.3, $\mathfrak{m}'(z) < 1$. Since $z_0 + u > 0$, therefore, we have

$$D'(u) = \mathfrak{m}'(z_0+u) - 2 < 1 - 2 < 0.$$

Since $D(0) = \mathfrak{m}(z_0) > 0$, and $\mathfrak{m}(z) \sim z$, it implies that $\lim\limits_{u\to+\infty} D(u) = \lim\limits_{u\to+\infty} -u = -\infty$. We conclude that $D(u)$ has a unique positive root, $\bar{u}_{t,g}$ that uniquely solves Eq. (B.2) and

$$\bar{\alpha}_{t,g} = -\frac{\bar{u}_{t,g}\sigma_0}{\sqrt{1-x_t^{(n+t-1)}}}.$$

In the case of the segment-based search, we have $\Delta = 0$. Following the identical analysis, we conclude that $\bar{u}_{t,s}$ uniquely solves

$$\mathfrak{m}(u) - 2u = 0,$$

and

$$\bar{\alpha}_{t,s} = -\frac{\bar{u}_{t,s}\sigma_0}{\sqrt{1-x_t^{(n+t-1)}}}.$$

$\square$

C Helpful Results

Lemma C.1. *H(z) is convex and decreasing in z, and* $\lim_{z\to+\infty} H(z) = 0$*, for* $z \in \mathbb{R}$*.*

Proof of Lemma C.1. Recall the definition of $H(\cdot)$ in (4.4):

$$H(z) = \phi(z) - z\left(1 - \Phi(z)\right).$$

Therefore,

$$H'(z) = -z\phi(z) - (1 - \Phi(z)) + z\phi(z) = -\left(1 - \Phi(z)\right) \le 0.$$

Further,

$$H''(z) = \phi(z) > 0.$$

For $\lim_{z\to+\infty} H(z)$, we have

$$\begin{aligned}
\lim_{z\to+\infty} H(z) &= \lim_{z\to+\infty} \left[\phi(z) - z\left(1 - \Phi(z)\right)\right] \\
&= -\lim_{z\to+\infty} \frac{1 - \Phi(z)}{\frac{1}{z}} \qquad \left(\text{since } \lim_{z\to+\infty} \phi(z) = 0\right) \\
&= -\lim_{z\to+\infty} \frac{\phi(z)}{\frac{1}{z^2}} \qquad (\text{by L'Hôpital's rule}) \\
&= -\lim_{z\to+\infty} \frac{1}{\sqrt{2\pi}} \frac{z^2}{e^{\frac{z^2}{2}}} \\
&= -\lim_{t\to+\infty} \sqrt{\frac{2}{\pi}} \frac{t}{e^t} \qquad \left(\text{where } t = \frac{z^2}{2}\right) \\
&= -\lim_{t\to+\infty} \sqrt{\frac{2}{\pi}} \frac{1}{e^t} = 0 \qquad (\text{by L'Hôpital's rule}).
\end{aligned}$$

□

Lemma C.2. *[Gordon 1941; Mitrinovic and Vasic 1970; Baricz 2010; Wainwright 2019] Define* $\varphi(z) = \frac{1-\Phi(z)}{\phi(z)}$ *as the Mills-Ratio of standard normal distribution. We have*

- $\varphi'(z) = \varphi(z) - 1 < 0, \quad z \in \mathbb{R}$
- $\varphi''(z) = \varphi(z) + z\varphi'(z) > 0, \quad z \in \mathbb{R}$
- $\varphi(z) < \frac{1}{z}, \quad z > 0,$
- $0 < \frac{1}{\varphi(z)} - z < \frac{1}{z}, \quad z > 0.$

Lemma C.3 (Sampford (1953))**.** *Let* $\mathfrak{m}(z) = \frac{1}{\varphi(z)}$ *be the inverse Mills-Ratio. We have*

$$0 < \mathfrak{m}'(z) < 1.$$

Lemma C.4. *Let* $T \in \mathbb{N}$*. If* $f : [1, T] \to \mathbb{R}$ *is a differentiable and increasing function. Then,*

$$\int_1^T f(x)dx + f(1) \le \sum_{t=1}^T f(t) \le \int_1^T f(x)dx + f(T).$$

Proof of Lemma C.4. For any $t \in [T-1], x \in [t, t+1]$, we have

$$f(t) \leq f(x) \leq f(t+1).$$

Integrating over $[t, t+1]$ gives

$$f(t) \leq \int_t^{t+1} f(x)dx \leq f(t+1).$$

Summing these inequalities for $t \in [T-1]$ yields

$$\sum_{t \in [T-1]} f(t) \leq \int_1^T f(x)dx \leq \sum_{t=2}^T f(t).$$

Adding $f(1)$ to the middle and right terms, and adding $f(T)$ to the middle and left terms, we conclude the proof. $\square$

Lemma C.5. *$H\left(\frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}\right)$ is unimodal in $x \in \mathcal{S}_t^{(i)}$; $\lim_{x \downarrow x_t^{(i)}} g_t'(x) \geq 0$; $\lim_{x \uparrow x_t^{(i+1)}} g_t'(x) \leq 0$.*

Proof of Lemma C.5. Since H is a decreasing function, it is sufficient to show that $z_t(x) = \frac{f_t^\star - \mu_t(x)}{\sigma_t(x)}$ has a unique minimizer in $\mathcal{S}_t^{(i)}$.

First, we perform the following variable transformation.

$$\begin{aligned}
\Delta_t^{(i)} &= f_t^\star - f\left(x_t^{(i)}\right) \geq 0, \quad & \Delta_t^{(i+1)} &= f_t^\star - f\left(x_t^{(i+1)}\right) \geq 0, \\
a_t^{(i)} &= x_t^{(i+1)} - x_t^{(i)} > 0, \quad & k_t^{(i)} &= \Delta_t^{(i)} - \Delta_t^{(i+1)} = f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right).
\end{aligned}$$

Let

$$s = \frac{x - x_t^{(i)}}{a_t^{(i)}} \in [0,1] \Leftrightarrow x = (1-s)x_t^{(i)} + s x_t^{(i+1)} \in \left[x_t^{(i)}, x_t^{(i+1)}\right].$$

It follows that

$$z_t(s) = \frac{f_t^\star - \mu_t(s)}{\sigma_t(s)} = \frac{\Delta_t^{(i)} - k_t^{(i)} s}{\sigma_t(s)} \geq 0, \;\; \sigma_t(s) = \sigma_0 \sqrt{a_t^{(i)}} \sqrt{s(1-s)}.$$

The case where $k_t^{(i)} = 0$ is trivial. Consider the case where $k_t^{(i)} \neq 0$.

By the quotient rule, we have

$$z_t'(s) = \frac{-k_t^{(i)} \sigma_t(s) - \left(\Delta_t^{(i)} - k_t^{(i)} s\right) \sigma_t'(s)}{\sigma_t^2(s)} \tag{C.1}$$

$$\sigma_t'(x) = \sigma_0 \sqrt{a_t^{(i)}} \frac{1-2s}{2\sqrt{s(1-s)}} = \sigma_t(s) \frac{1-2s}{2s(1-s)} \tag{C.2}$$

Combine (C.2) and (C.1), giving us

$$\begin{aligned} z_t'(s) &= \frac{-k_t^{(i)}\sigma_t(s) - \left(\Delta_t^{(i)} - k_t^{(i)} s\right)\sigma_t(s)\frac{1-2s}{2s(1-s)}}{\sigma_t^2(s)} \\ &= \frac{\left(\Delta_t^{(i)} + \Delta_t^{(i+1)}\right)s - \Delta_t^{(i)}}{2\sigma_t(s)s(1-s)}. \end{aligned} \tag{C.3}$$

Since the denominator is strictly positive in $(0,1)$, the sign of $z_t'(s)$ is identical to the sign of the affine function in the numerator. The unique root of $z_t'(s)$ is

$$\bar{s}_t^{(i)} = \frac{\Delta_t^{(i)}}{\Delta_t^{(i)} + \Delta_t^{(i+1)}}.$$

As $f\left(x_t^{(i+1)}\right) > f\left(x_t^{(i)}\right) \Leftrightarrow \Delta_t^{(i)} > \Delta_t^{(i+1)}$, we conclude that $\bar{s}_t^{(i)} \in \left[\frac{1}{2}, 1\right]$.

Now, we prove that $\lim_{x\downarrow x_t^{(i)}} g_t'(x) \geq 0$ and $\lim_{x\uparrow x_t^{(i+1)}} g_t'(x) \leq 0$.

Without l.o.g., let $f\left(x_t^{(i+1)}\right) > f\left(x_t^{(i)}\right)$. In the space of s, we have

$$\begin{aligned} g_t(s) &= \sigma_t(s) H\left(\frac{f_t^\star - \mu_t(s)}{\sigma_t(s)}\right) \\ &= \sigma_t(s) H\left(\frac{\Delta_t^{(i)} - k_t^{(i)} s}{\sigma_t(s)}\right) \end{aligned} \tag{C.4}$$

Then, the derivative of $g_t(s)$ w.r.t. s is

$$g_t'(s) = k_t^{(i)}\left(1 - \Phi(z_t(s))\right) + \sigma_t'(s)\phi(z_t(s)) = \phi(z_t(s))\left[k_t^{(i)}\varphi\left(z_t(s)\right) + \sigma_t'(s)\right]. \tag{C.5}$$

It is trivial when $k_t^{(i)} = 0$. We focus on the case when $k_t^{(i)} \neq 0$.

Since $f\left(x_t^{(i+1)}\right) > f\left(x_t^{(i)}\right)$ implies that $\Delta_t^{(i)} > 0, \Leftrightarrow f_t^\star > f\left(x_t^{(i)}\right)$, and $\Delta_t^{(i)} \geq k_t^{(i)}$, we have

$$\lim_{s\downarrow 0} z_t(s) = \lim_{s\downarrow 0} \frac{\Delta_t^{(i)}}{\sigma_t'(s)} = +\infty,$$

and

$$z_t(s) = \frac{\Delta_t^{(i)} - k_t^{(i)} s}{\sigma_t(s)} \geq \frac{k_t^{(i)}}{\sigma_0\sqrt{a_t^{(i)}}}\sqrt{\frac{1}{s} - 1}$$

$$\sigma_t'(s) \leq \frac{\sigma_0\sqrt{a_t^{(i)}}}{2}\sqrt{\frac{1}{s} - 1}.$$

Now we analyze $\lim\limits_{x\downarrow 0} g_t'(s)$.

For the first term in $g_t'(s)$, we have:

$$\lim_{s\downarrow 0} k_t^{(i)}\left(1 - \Phi(z_t(s))\right) = \lim_{z\to+\infty} k_t^{(i)}\left(1 - \Phi(z)\right) = 0.$$

For the second term in $g_t'(s)$, we have:

$$0 \leq \sigma_t'(s)\phi(z_t(s)) \leq \frac{\sigma_0\sqrt{a_t^{(i)}}}{2}\sqrt{\frac{1}{s}-1} \times \phi\left(\frac{k_t^{(i)}}{\sigma_0\sqrt{a_t^{(i)}}}\sqrt{\frac{1}{s}-1}\right).$$

Since

$$\lim_{s\downarrow 0} \frac{\sigma_0\sqrt{a_t^{(i)}}}{2}\sqrt{\frac{1}{s}-1} \times \phi\left(\frac{k_t^{(i)}}{\sigma_0\sqrt{a_t^{(i)}}}\sqrt{\frac{1}{s}-1}\right) = \lim_{z\to+\infty} \frac{\sigma_0\sqrt{a_t^{(i)}}}{2\sqrt{2\pi}}\sqrt{z}e^{-\frac{z}{2}} = 0,$$

we have $\lim_{x\downarrow 0} g_t'(s) = 0$ in this case.

As for the limit of $g_t'(s)$ at the right boundary, i.e., $\lim_{s\uparrow 1} g_t'(s)$, it is straightforward to verify the following

$$\lim_{s\uparrow 1} z_t(s) = \begin{cases} 0 & f_t^\star = f\left(x_t^{(i+1)}\right) \\ +\infty & f_t^\star > f\left(x_t^{(i+1)}\right) \end{cases} \Rightarrow \lim_{s\uparrow 1} g_t'(s) = \begin{cases} +\infty & f_t^\star = f\left(x_t^{(i+1)}\right) \\ 0 & f_t^\star > f\left(x_t^{(i+1)}\right) \end{cases}$$

This limit behavior is symmetric in the case of $f\left(x_t^{(i+1)}\right) < f\left(x_t^{(i)}\right)$.

$\square$

Lemma C.6. *$\frac{f_t^{\star(i)}-\mu_t(x)}{\sigma_t(x)}$ is increasing in x if and only if $f\left(x_t^{(i)}\right) \geq f\left(x_t^{(i+1)}\right)$.*

Proof of Lemma C.6. In the segment-based search, $f_t^{\star(i)} = \max\left\{f\left(x_t^{(i+1)}\right), f\left(x_t^{(i)}\right)\right\}$. Analogous to the variable transformation as presented in the proof of Lemma C.5, we have

$$\begin{aligned} z_t'(s) &= \frac{-k_t^{(i)}\sigma_t(s) - \left(\Delta_t^{(i)} - k_t^{(i)}s\right)\sigma_t(s)\frac{1-2s}{2s(1-s)}}{\sigma_t^2(s)} \\ &= \frac{\left(\Delta_t^{(i)} + \Delta_t^{(i+1)}\right)s - \Delta_t^{(i)}}{2\sigma_t(s)s(1-s)} \end{aligned}$$

where

$$\begin{aligned} &\Delta_t^{(i)} = f_t^{\star(i)} - f\left(x_t^{(i)}\right) \geq 0, \quad && \Delta_t^{(i+1)} = f_t^{\star(i)} - f\left(x_t^{(i+1)}\right) \geq 0, \\ &a_t^{(i)} = x_t^{(i+1)} - x_t^{(i)} > 0, && k_t^{(i)} = \Delta_t^{(i)} - \Delta_t^{(i+1)} = f\left(x_t^{(i+1)}\right) - f\left(x_t^{(i)}\right). \end{aligned}$$

In the case of $f_t^{\star(i)} = f\left(x_t^{(i+1)}\right)$, $z_t'(s) = -\frac{\Delta_t^{(i)}}{2\sigma_t(s)s} \leq 0$; otherwise, $z_t'(s) = \frac{\Delta_t^{(i+1)}}{2\sigma_t(s)(1-s)} \geq 0$.

$\square$