Saving Kermit: Dynamic Assortment Planning in a Boiling Market

Eneko C. Clemente[1], Oleg A. Prokopyev[1], and Denis Sauré[2]

[1]Department of Business Administration, University of Zurich

[2]Industrial Engineering Department, University of Chile

Problem definition. A retailer planning its assortment under evolving customers' preferences faces a tension between retaining past sales information to guide its assortment selection (*remembering*) and discarding that information once it no longer reflects current purchasing behavior (*forgetting*). Reminiscent of the boiling frog apologue, ignoring this tension may quietly erode the retailer's revenue as the offered assortment becomes misaligned with customers' evolving preferences. ***Methodology and results.*** We treat evolving customers' preferences as an operational risk that the retailer plans for upfront. The retailer chooses its *risk profile* by specifying a class of admissible preferences' dynamics against which robustness is sought. Given that class, we formulate robust planning as minimizing regret relative to a clairvoyant retailer that knows how preferences evolve. We show that restart-based strategies achieve robustness when the restart timing is tied to the retailer's risk profile. ***Managerial implications.*** To manage the risk of evolving customers' preferences, a retailer should first articulate a risk profile aligned with the dynamics it anticipates and then use that profile to determine the restart timing. Broad robustness is costly, whereas focusing on structured dynamics moves robust planning from pre-planned restarts to adaptive ones triggered only when preferences change.

1 Introduction

The growth of e-commerce and data-intensive retailing has made assortment planning a central operational decision for modern retailers. Digital channels allow firms to rapidly adjust the set of products they offer and to observe customers' choices at an unprecedented level of granularity. These capabilities create new opportunities to infer customers' preferences and tailor assortments accordingly (Bernstein et al. 2019; Caro et al. 2020). At the same time, retailers face a binding constraint: only a limited number of products can be displayed on prominent pages such as homepages and search results. This constraint forces retailers to be highly selective, making the alignment between displayed assortments and customers' preferences a key determinant of sales performance.

The dynamic nature of customers' preferences has been recognized since the 1960s, with early contributions from behavioral learning theory shaping our understanding of brand loyalty (J. N. Sheth 1967). Demographic shifts and the gradual evolution of long-term tastes can subtly yet persistently alter customers' behavior (Döpper et al. 2024). These dynamics are particularly consequential in assortment planning. Because assortment decisions are combinatorial, even small changes in customers' preferences can lead to large shifts in which assortment is most profitable.

While some shifts in preferences unfold gradually and may be largely irrelevant over the short lifecycle of a product, others occur as sudden shocks. Abrupt disruptions, such as financial crises or pandemics, can overturn purchasing patterns within weeks, forcing retailers to adjust their assortments rapidly. Recessions, for example, tend to shift demand toward budget-friendly alternatives, disadvantaging premium products (Hampson and McGoldrick 2013). Similarly, pandemics can generate sudden spikes in demand for essential goods, such as medical supplies, while sanitary restrictions sharply depress the consumption of services (J. Sheth 2020).

Changes in customers' preferences not only shape the environment in which retailers operate, but also create opportunities to better understand market dynamics (Hartmann and Nair 2010). Retailers can leverage these changes when making core operational decisions, from pricing to assortment planning (Rooderkerk et al. 2022). Yet, much of the dynamic assortment planning literature continues to assume static customers' preferences, thereby understating the risks associated with preferences' evolution. This risk is well illustrated by the *boiling frog* apologue: gradual changes may go unnoticed until their consequences become severe. Likewise, a retailer that relies on a fixed assortment plan may be caught unprepared as preferences evolve over time.

Responding to evolving preferences from sales data alone is challenging and often requires sophisticated algorithms (Foussoul et al. 2023). Meanwhile, such flexible responses introduce operational uncertainty, since both the timing and the trajectory of preferences' changes are typically unknown. Retailers can mitigate this uncertainty by planning for this risk upfront through robust decision-making. Such planning requires specifying, in advance, the class of preferences' dynamics against which protection is sought. To discipline this choice, retailers can draw on *frontline inputs* such as employee knowledge (Ton 2014) and external signals such as consumer indices and social media sentiment (Caro et al. 2020; Rooderkerk et al. 2022). Together, these inputs help define a retailer's *risk profile*: how conservative it chooses to be and which preferences' dynamics it seeks to hedge against.

Objective. We study dynamic assortment planning with evolving customers' preferences. Our goal is to inform retailers about how their risk profile affects their assortment planning, both operationally and in terms of revenue, when preferences might evolve over time. In such environments, retailers can adapt existing policies that account for evolving preferences, for example by periodically restarting the learning process. This adaptation gives rise to a fundamental trade-off between *remembering* past information about preferences and *forgetting*, that is, restarting learning when historical data may no longer reflect customers' behavior. We characterize how the retailer's risk profile governs the balance of this trade-off, and, in particular, how it determines the timing of such restarts.

Model. We study a dynamic assortment planning problem over a known horizon T, where utility-maximizing customers arrive sequentially and decide whether to purchase from an offered assortment. The retailer aims to maximize profit while facing uncertainty about how customers' preferences evolve over time. Our problem belongs to the domain of sequential decision-making under uncertainty (Hannan 1957). To address the challenge posed by unknown and evolving preferences, we study assortment strategies that are robust to a broad class of changes in customers' behavior.

Defining the retailer's risk profile requires specifying a class of admissible preferences' dynamics against which robustness is sought. Frontline inputs can discipline this choice in several ways, for instance by restricting attention to multinomial logit preferences with bounded attraction parameters, by ruling out joint changes for certain product pairs, or by focusing on large but infrequent shifts when product lifecycles are short. To accommodate a broad range of risk profiles, we characterize the class of admissible preferences' dynamics along three dimensions that are central to assortment planning: (*i*) *velocity*, capturing how quickly preferences evolve, (*ii*) *magnitude*, capturing the extend of preferences' changes; and (*iii*) *detectability*, capturing whether changes can be inferred from sales data under the assortments currently offered by the retailer. Changes are *passively* detectable when they occur within the retailer's current assortment and can be inferred from observed purchases, and *actively* detectable when they can be revealed only by offering alternative assortments.

We evaluate the retailer's performance by comparing the expected revenue generated under a given *policy* (i.e., an assortment strategy) to that of a clairvoyant retailer endowed with perfect knowledge of customers' preferences. The resulting difference, known as *regret*, is a standard performance metric in the online learning literature (Foster and Vohra 1999). Regret captures the *opportunity cost* associated with operating under incomplete information about preferences. To account for the revenue risk induced by the unknown and evolving preferences, we formulate the retailer's objective as minimizing worst-case regret and study how this quantity scales with the time horizon T. This objective admits a game-theoretic interpretation: the retailer commits to an assortment policy, after which the environment responds adversarially by selecting a sequence of customers' preferences, subject to the constraints imposed by the retailer's risk profile.

Under static preferences, dynamic assortment planning typically focuses on the classical trade-off between *exploration* and *exploitation* (Lai and Robbins 1985), balancing the need to explore different assortments to collect data on customers' purchasing behavior against the goal of exploiting current estimates to maximize immediate revenue (Caro and Gallien 2007). We argue that this approach is insufficient when customers' preferences evolve over time. In such settings, a distinct trade-off between *remembering*, that is, relying on historical data, and *forgetting*, namely discarding data that no longer reflects current customers' preferences, becomes central. We examine this trade-off, and show how it is shaped by the three dimensions that characterize the retailer's risk profile.

Velocity	Magnitude	Detectability	Planning regime	Regret (informal)	Section
-	M_T	-	Pre-planned restarts	$\mathcal{O}\big(\mathcal{R}^{\mathcal{A}}\big(T^{\frac{1}{2}}M_T^{-\frac{1}{2}}\big)\cdot T^{\frac{1}{2}}M_T^{\frac{1}{2}}\big)$	4
Abrupt	bounded	Active	Active monitoring	$\mathcal{O}\big(\sqrt{T}\log T+\mathcal{R}^{\mathcal{A}}(T)\big)$	5.2
Abrupt	bounded	Passive	Passive monitoring	$\mathcal{O}\big(\log T+\mathcal{R}^{\mathcal{A}}(T)\big)$	5.3

Table 1: Summary of our three key contributions. The retailer's risk profile is characterized along three dimensions of the class of admissible preferences' dynamics against which robustness is sought: velocity, magnitude (denoted by M_T), and detectability. These dimensions determine both the planning regime through which robustness is achieved and the attainable regret. Here, $\mathcal{R}^{\mathcal{A}}(T)$ is the regret of an assortment strategy $\mathcal{A}$ over T customers with static preferences.

Contributions. We show that retailers can manage the risk of evolving preferences while exploiting dynamic assortment planning policies originally designed for static settings. Robust planning in this context begins by specifying *a priori* a class of admissible preferences' dynamics against which protection is sought. The characteristics of this class (i.e., its velocity, magnitude, and detectability) jointly determine the achievable regret and, consequently, the appropriate robust planning strategy. Aside from requiring preferences to follow a random-utility model, we impose no restrictions on the admissible dynamics. Our contributions, summarized in Table 1, are threefold.

First, we quantify the volatility in the class of admissible preferences' dynamics induced by the retailer's risk profile through a magnitude measure M_T, which captures how much preferences can move over the selling horizon. We establish a fundamental performance limit showing that no assortment strategy can achieve regret below $\mathcal{O}\big(T^{3/4}M_T^{1/4}\big)$. This bound reveals an intrinsic jump in the opportunity cost relative to dynamic assortment planning with static preferences, where worst-case regret is typically of order $\mathcal{O}\big(\sqrt{T\log T}\big)$; see Agrawal et al. (2019).

To obtain a matching upper bound on the regret, we adopt a restart-based strategy that partitions the horizon into epochs and, within each epoch, applies a dynamic assortment planning algorithm designed for static preferences. Restart-based strategies are standard in the literature (Besbes et al. 2015), and their simplicity makes them particularly well suited to our setting. They provide a transparent mechanism for translating the retailer's risk profile into a predictable policy and for exposing the underlying remembering–forgetting trade-off. We show that choosing the restart timing as a function of the retailer's risk profile achieves the lower bound on the regret up to constant factors. As a result, improvements in dynamic assortment planning under static preferences carry over directly to the setting with evolving preferences through this restart mechanism.

Second, we show that not all restrictions on the class of admissible preferences' dynamics are equally valuable and that the retailer can trade some robustness for exploitable structure. Motivated by rare but impactful shocks such as financial crises or pandemics, we restrict the class of admissible preferences' dynamics to single abrupt shifts from *pre-change* to *post-change* preferences, with both the timing and the size of the change unknown *a priori*. From an operational perspective, a retailer may not wish to hedge against arbitrarily small abrupt shifts, since reacting to minor fluctuations can create instability and therefore unpredictability. We capture this risk profile by imposing a lower bound on the size of the change. In this high-velocity regime, the primary challenge shifts from managing volatility to detecting when a change in preferences occurs before it gradually erodes revenue.

In this setting, robust planning becomes a change-detection problem, a classical topic in statistics (Lorden 1971). We integrate these detection ideas into dynamic assortment planning and show that robust planning shifts from pre-scheduled restarts to proactive monitoring. This approach yields regret of order $\mathcal{O}(\sqrt{T})$, substantially smaller than the $\mathcal{O}(T^{3/4})$ rate in the general case.

Third, we distinguish two regimes based on *detectability*. A change is *passively detectable* if it affects products already included in the pre-change optimal assortment, so that monitoring sales from that assortment suffices to reveal the shift. By contrast, a change is *actively detectable* if it affects products outside that assortment, in which case detection requires deliberate experimentation with alternative assortments. We show that detectability has first-order performance implications. When changes are passively detectable, the regret's order is only $\mathcal{O}\big(\log T\big)$ on top of learning post-change preferences, whereas actively detectable changes lead to regret of order $\mathcal{O}\big(\sqrt{T}\big)$. Thus, evolving preferences need not affect all retailers equally, even under a similar risk profile. A retailer's assortment may be sufficiently representative of its broader product range for changes to be passively detectable.

Managerial insights. Evolving preferences, much like in the boiling frog apologue, can quietly erode revenue when retailers rely on policies designed for static settings and fail to manage the remembering–forgetting trade-off. Effective risk management therefore begins with upfront planning, by articulating a risk profile that specifies the preferences' dynamics against which robustness is sought. Restart-based strategies, with timing tied to this risk profile, offer a practical lever for achieving robustness. Robustness to arbitrary dynamics is costly, whereas trading some robustness for exploitable structure (for example, hedging only against abrupt shifts) can sharpen the restart timing and materially improve performance. We illustrate these insights through a case study using data from a Chilean retailer, which highlights the cost of failing to plan for evolving preferences, the gains from tightening the risk profile, and the performance loss from excessive conservatism.

Organization. Section 2 reviews the literature. Section 3 formulates the dynamic assortment planning problem with evolving customers' preferences. Section 4 discusses a general principle to handle evolving preferences. Section 5 studies how the retailer's risk profile shape robust planning. Section 6 presents a case study using data from a Chilean retailer. Finally, Section 7 concludes.

2 Literature review

Consumer's behavior. Discrete choice models have become central in assortment planning (Kök et al. 2015), following early work by Mahajan and Van Ryzin (2001) and Talluri and Van Ryzin (2004). Most of the choice models are utility-based and represent preferences by mapping the utility that customers derive from each product to its purchase probability (Train 2009). Initial studies focus on parametric approaches, particularly the multinomial logit (MNL) model, valued for its analytical simplicity (Rusmevichientong et al. 2010). Yet, the MNL model's "independence of irrelevant alternatives" property limits its ability to capture realistic substitution patterns.

Extensions such as the mixed logit (Feldman and Topaloglu 2015), nested logit (Gallego and Topaloglu 2014), and Markov chain choice (MCC) models (Blanchet et al. 2016) capture more realistic substitution patterns but introduce new challenges. The mixed logit may overfit in data-scarce settings, while MCC lacks a closed-form expression for purchase probabilities which complicates its estimation. In parallel, non-parametric approaches based on consumers' product rankings have also attracted considerable attention (Honhon et al. 2012; Van Ryzin and Vulcano 2015).

In our study, we represent customers' preferences with random-utility choice models. To keep the approach flexible and let retailers accommodate rich behavioral effects, we do not restrict attention to a particular parametric or non-parametric model. Moreover, to capture evolving substitution patterns, we allow the underlying choice model itself to vary over the selling horizon.

Role of learning. Since customers' preferences are typically unknown by the retailer, studies have emerged on "learning" them from the data. Learning customers' preferences can be broadly divided into two settings. In the *offline* setting, pre-existing sales data are used to calibrate a model for customers' preferences. For example, Farias et al. (2013) introduce a data-driven method that relaxes traditional parametric assumptions by inferring the model's structure directly from sales data. In the *online* setting, the retailer faces a sequential decision-making under uncertainty problem. It learns by continuously updating its estimate of preferences "on the fly" from both the observed sales data and the assortments displayed to the customers. In this study, we focus on the online setting.

Multi-armed bandit (MAB) algorithms are commonly used in online learning to balance exploration, gathering information about customers' preferences, and exploitation, maximizing immediate revenue (Cesa-Bianchi and Lugosi 2006). Building on this foundation, the seminal work by Caro and Gallien (2007) introduces an MAB approach for dynamic assortment planning. Subsequent contributions by Rusmevichientong et al. (2010) and Sauré and Zeevi (2013) incorporate choice models into the bandit setting and achieve regret of order $\mathcal{O}(\log T)$, matching the asymptotic lower bound established by Lai and Robbins (1985). Moreover, these techniques have been refined for various choice models, such as for MNL (Agrawal et al. 2019) and for MCC (Li et al. 2025).

A large body of the literature relies on the assumption that customers' preferences remain static over the selling horizon. Given the maturity of this setting, we leverage these advances by adapting existing assortment planning policies to environments with evolving preferences through restart-based strategies. The restart timing is treated as a central operational decision, since it determines when the retailer re-estimates preferences and redeploys a policy as if operating in a new regime.

Learning in varying environments. Customers' preferences need not be static, yet many assortment models in the literature assume otherwise. The related problem of learning demand in dynamic pricing is studied by Besbes and Zeevi (2009) and extended to changing demand by Besbes and Zeevi (2011), Besbes and Sauré (2014), and Keskin and Zeevi (2017). The pricing setting differs from assortment planning in two ways. First, small shifts in demand often lead to small changes in the optimal price, whereas even minor shifts in preferences can trigger large changes in the optimal assortment. Second, products that are not offered generate no sales data, so preferences for these products cannot be learned without deliberately offering them. Closer to our assortment planning problem, Foussoul et al. (2023) study dynamic assortment planning under a time-varying MNL model. Their study builds on the sophisticated reduction by Wei and Luo (2021) which extends stationary online-learning algorithms to non-stationary environments with minimal assumptions.

In contrast to Foussoul et al. (2023), we view evolving preferences as an operational risk to be managed, rather than as a non-stationary learning problem in which adaptation is the primary objective. We let the retailer define a risk profile using available frontline inputs by specifying a class of admissible preferences' dynamics against which robustness is sought. *Our contribution is therefore operational rather than algorithmic:* we translate the retailer's risk profile into a predictable restart rule for the estimation of preferences. Because our approach treats a dynamic assortment planning policies designed for static preferences as subroutines, any algorithmic improvement in that setting caries over directly into revenue gains in our framework. Ultimately, the retailer's decision reduces to choosing how much risk to hedge against and using that choice to determine the restart timing.

3 Problem formulation

Model primitives. We consider an assortment planning problem for a retailer offering $N \in \mathbb{N}$ differentiated products. Each product $i \in \mathcal{N} := [N]$ is sold at a price $r_i > 0$, yielding a profit $w_i \equiv r_i - c_i > 0$, where c_i denotes the marginal acquisition cost. Customers arrive sequentially (one per period) over a known horizon, with each customer indexed by some $t \in [T]$. This horizon is determined by the number of arrivals, so we use "time" and "customer" interchangeably throughout the paper. Each customer $t \in [T]$ assigns a random utility U_i^t to each product $i \in \mathcal{N}_0 := \mathcal{N} \cup \{0\}$, with $i = 0$ representing the no-purchase option. The joint distribution of these utilities, denoted by F^t, characterizes the preferences of customer t. We impose *no* additional structure on $F^{(\mathbb{N})} \equiv (F^t : t \in \mathbb{N})$ beyond requiring that they share a common probability space and satisfy:

$$\mathbb{P}(U_i^t = U_j^t) = 0, \quad \forall\, i, j \in \mathcal{N}_0,\, i \neq j,\, \forall\, t \in \mathbb{N}.$$

Upon arrival, each customer $t \in [T]$ is presented with an assortment S^t chosen from a set $\mathcal{S}$ of product mixes of size at most K, defined by $\mathcal{S} := \{S \subseteq \mathcal{N} : |S| \leq K\}$. Given this assortment, the customer then makes a purchase decision that maximizes its intrinsic utility. That is,

$$i_t \in \operatorname{argmax} \left\{ U_i^t \,:\, i \in S^t \cup \{0\} \right\}$$

denotes the purchase decision of customer $t \in [T]$.

Single-sale assortment planning. We assume that the retailer faces neither inventory constraints nor switching costs. These assumptions are commonly adopted in the dynamic assortment planning literature (Sauré and Zeevi 2013; Agrawal et al. 2019; Li et al. 2025) to isolate the cost of learning customers' preferences in terms of revenue. We let $r(S^t, F^t)$ denote the expected revenue associated with offering assortment S^t to customer t. Formally, it is defined as follows:

$$r(S^t, F^t) := \sum_{i \in S^t} w_i\, p_i(S^t, F^t),$$

where $p_i(S^t, F^t)$ denotes the probability that customer t buys product $i \in \mathcal{N}_0$ within the displayed assortment $S^t \in \mathcal{S}$, when utilities are distributed according to F^t.

Specifically, we define the purchasing probability for each product $i \in S^t \cup \{0\}$ from the offered assortment, including the no-purchase option, as:

$$p_i(S^t, F^t) := F^t\big(\{x \in \mathbb{R}^{|\mathcal{N}_0|} : x_i \geq x_j \text{ for all } j \in S^t \cup \{0\}\}\big).$$

For products that are not included in the assortment, we set the purchase probability to zero, i.e., $p_i(S^t, F^t) = 0$ for all $i \notin S^t$. For each customer t, we define the single-sale *optimal assortment* as:

$$S^*(F^t) \in \operatorname{argmax} \left\{ r(S, F^t) \,:\, S \in \mathcal{S} \right\}.$$

Throughout the paper, we assume that the single-sale optimal assortment is uniquely defined and denote it by $S_t^* \equiv S^*(F^t)$ to isolate the effect of evolving preferences on assortment decisions.

Retailer's risk profile. Unlike traditional dynamic assortment planning models, we assume that customers' preferences, represented by a sequence of preferences $F^{(\mathbb{N})}$, evolve over time rather than remaining static, i.e., $F^t \equiv F$ for all $t \in \mathbb{N}$ and a fixed distribution F. Because preferences are not directly observable, the sequence $F^{(\mathbb{N})}$ is unknown *a priori*. However, retailers rarely operate in a vacuum. They can often rely on frontline inputs to identify how preferences tend to move. For example, store staff who interact with customers on a day-to-day basis can identify close substitutes, and customer service teams can observe recurring requests for specific features.

The retailer leverages such information to construct a *class* $\mathcal{F}$ of preferences' dynamics against which it seeks robustness. It does so by imposing restrictions such as a parametric choice model with parameters constrained to lie in a specified set, a predefined substitution structure across products, and bounds on how many changes in preferences can occur and how large each change can be. The restrictions embedded in $\mathcal{F}$ encode the retailer's *risk profile*. Furthermore, we assume that $F^{(\mathbb{N})} \in \mathcal{F}$ is such that $p_i(S, F^t) \in (0, 1)$ for all $i \in S$, $S \in \mathcal{S}$, and $t \in \mathbb{N}$. This technical condition excludes degenerate cases in which customers make deterministic choices.

Planning under evolving preferences. Let $\mathcal{H}_t := \sigma\big((S^s, i_s) \,:\, s < t\big)$ denote the history of offered assortments and purchases prior to customer $t \in [T]$. A sequence of (possibly random) assortments $\pi := \big(S_t^\pi \,:\, t \leq T\big)$ is called an *admissible policy* if at each time t, the assortment decision S_t^π is a mapping that takes as an input the past history $\big(\mathcal{H}^s \,:\, s < t\big)$ for all $t \in [T]$ and returns a feasible assortment from $\mathcal{S}$. Specifically, S_t^π represents the assortment offered by policy π to customer t. We denote by $\mathcal{P}$ the set of all such assortment strategies.

Following the literature (Cesa-Bianchi and Lugosi 2006), we measure the performance of any assortment strategy against that achieved by a clairvoyant retailer (called *oracle*) with prior knowledge on $F^{(\mathbb{N})}$. Specifically, this oracle knows the preferences F^t of customer t, and therefore offers assortment S_t^* to said customer. For given preferences $F^{(\mathbb{N})} \in \mathcal{F}$, we define the *oracle revenue* as:

$$J^*\big(F^{(\mathbb{N})}, T\big) := \sum_{t \leq T} r\left(S_t^*, F^t\right).$$

In essence, the oracle revenue represents the best achievable profit and is not attainable in practice, since the retailer lacks perfect knowledge of customers' preferences. In particular, an assortment strategy $\pi \in \mathcal{P}$ achieves in expectation a cumulative revenue given by:

$$J^{\pi}\big(F^{(\mathbb{N})}, T\big) := \mathbb{E}\big\{\sum_{t \le T} r(S_t^{\pi}, F^t)\big\},$$

where the expectation is taken over the (random) assortments $\big(S_t^{\pi} : t \le T\big)$ offered by strategy π.

Because the retailer does not know in advance which sequence of preferences from $\mathcal{F}$ will be realized, we adopt a robust planning perspective and evaluate performance relative to a clairvoyant retailer under an adversarial realization of $F^{(\mathbb{N})}$. Importantly, the adversary is not unconstrained. It must select $F^{(\mathbb{N})}$ within the class $\mathcal{F}$, which encodes the retailer's risk profile. Consequently, weaker restrictions on $\mathcal{F}$ enlarge the adversary's choice set, while stronger restrictions narrow it. We define the *adversarial regret* of an assortment strategy $\pi \in \mathcal{P}$ as:

$$\mathcal{R}^{\pi}\big(\mathcal{F}, T\big) := \sup\big\{J^{*}\big(F^{(\mathbb{N})}, T\big) - J^{\pi}\big(F^{(\mathbb{N})}, T\big) \ : \ F^{(\mathbb{N})} \in \mathcal{F}\big\}.$$

The adversarial regret quantifies the worst-case opportunity cost incurred by the retailer when committing to an assortment strategy under the risk profile encoded by $\mathcal{F}$. It therefore captures the *price of robustness* to the class of admissible preferences' dynamics $\mathcal{F}$. The goal of robust planning is to construct an assortment strategy that minimizes this price. Formally, we define:

$$\mathcal{R}^{*}(\mathcal{F}, T) := \inf\left\{\mathcal{R}^{\pi}(\mathcal{F}, T) : \pi \in \mathcal{P}\right\},$$

as the lowest regret attainable by the retailer under the chosen risk profile.

4 Robust planning under evolving customers' preferences

To shed light on the risk of evolving preferences, we characterize the retailer's risk profile through key dimensions that enable us to derive a fundamental performance limit that no strategy can surpass. We show that this limit can be attained by a simple restart-based strategy when the restart timing is aligned with the retailer's risk profile. This result prescribes a predictable path to optimal performance and quantifies the unavoidable cost of robustness to evolving preferences.

4.1 A macro view of preferences' dynamics

We capture the retailer's risk profile at a macro level by quantifying how preferences can move over time. We introduce the *magnitude* and the *velocity* of the class $\mathcal{F}$ to describe how much preferences can change and how quickly within that class. We measure these changes using the Kullback-Leibler (KL) divergence (Thomas and Joy 2006). The KL divergence applies to any random-utility choice model through differences in assortment-level purchase probabilities and connects changes in preferences to what the retailer can learn from sales. Indeed, the KL divergence has an information-theoretic interpretation, as it determines how many purchases are needed to reliably detect a change.

Magnitude. For any assortment $S \in \mathcal{S}$ and time period $t > 1$, we denote by $\mathcal{K}^t(S)$ the KL divergence between consecutive preferences F^{t-1} and F^t whenever S is offered. For $T \in \mathbb{N}$, we define the *magnitude* $\mathcal{M}(\mathcal{F}, T)$ of the *class* $\mathcal{F}$ as the highest cumulative variation along a sequence of customers' preferences among all possible such sequences. That is:

$$\mathcal{M}(\mathcal{F}, T) := \sup \Big\{ \sum_{t=2}^{T} \max \big\{ \mathcal{K}^t(S) \, : \, S \in \mathcal{S} \big\} \, : \, F^{(\mathbb{N})} \in \mathcal{F} \Big\}.$$

Restricting the class $\mathcal{F}$ can reduce the corresponding magnitude $\mathcal{M}(\mathcal{F}, T)$. A small magnitude ensures that the volatility of preferences within $\mathcal{F}$ remains low. Importantly $\mathcal{M}(\mathcal{F}, T)$ is deliberately model-agnostic and quantifies volatility without relying on any assumptions about the class $\mathcal{F}$.

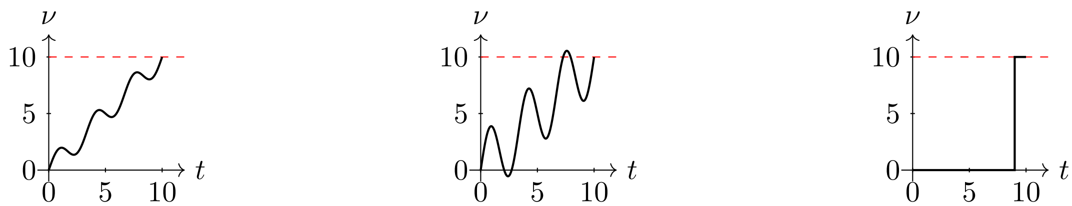
Example 1. We consider a retailer with $N = 10$ products, offering up to $K = 4$ items at any time. Each product i is assumed to yield a profit of $w_i = 1$. In this model, the utility of customer t for product $i \in \mathcal{N}_0$ (including the no-purchase option) is assumed to be given by $U_i^t = \mu_i^t + \varepsilon_i^t$, where μ_i^t represents the deterministic component of the utility, and ε_i^t is an idiosyncratic shock following a Gumbel distribution with location 0 and scale 1. Given an assortment S, the probability that a customer selects product $i \in S$ is given by $p_i(S, F^t) := \nu_i^t \cdot \left(\nu_0^t + \sum_{j \in S} \nu_j^t \right)^{-1}$, where $\left(\nu_i^t := \exp(\mu_i^t) \, : \, i \in [N] \cup \{0\} \right)$ are referred to as *attraction parameters* (Train 2009).

We define $\mathcal{F}_{\text{MNL}} \equiv \mathcal{F}_{\text{MNL}}(M_T)$, parameterized by $M_T > 0$, as the set of time-varying MNL models in which the attraction parameters ν_i^t switch between two regimes with equal probability at fixed intervals of length $\Delta = \lceil T^{1/2}(8M_T)^{-1/2} \rceil$. Specifically, for $1 \leq j \leq \lceil T/\Delta \rceil - 1$, we define ν_i^t such that for $t \in [(j-1)\Delta, j\Delta]$, $\nu_i^t = \nu_i^a$ with probability 1/2, and $\nu_i^t = \nu_i^b$ otherwise; the attraction parameters for products $i \in \{1, 2, 3, 4\}$ are $\nu_i^a = 0.25 + \zeta$, and for $i \in \{5, 6, 7, 8\}$ are $\nu_i^b = 0.25 + \zeta$, with $\zeta = \sqrt{M_T \Delta / T}$. For the other products, we set $\nu_i^a = \nu_i^b = 0.25$ and $\nu_0^a = \nu_0^b = 1$. Thus, one can verify that the optimal assortment for a single sale under ν^a is $S^*(\nu^a) = \{1, 2, 3, 4\}$, whereas it becomes $S^*(\nu^b) = \{5, 6, 7, 8\}$ under ν^b. Also, the parameter M_T, which may depend on T, drives the magnitude of the class $\mathcal{F}_{\text{MNL}}$ as one can verify that $\mathcal{F}_{\text{MNL}}$ satisfies $2M_T \leq \mathcal{M}(\mathcal{F}_{\text{MNL}}, T) \leq 8M_T$. ■

Velocity. Preferences may evolve gradually over time, exhibiting only marginal shifts from one customer to the next as opposed to abrupt disruptions. We refer to such scenarios as *slowly* changing preferences, distinguishing them from more rapid or sudden transitions. Formally, for $T \in \mathbb{N}$, we define the *velocity* of the class $\mathcal{F}$ as the maximum difference in preferences between consecutive customers across all possible sequences of preferences. That is:

$$\mathcal{V}(\mathcal{F},T) \coloneqq \sup\Big\{ \max\big\{\mathcal{K}^t(S) :\ S \in \mathcal{S},\ t = 2,\dots,T\big\} : F^{(\mathbb{N})} \in \mathcal{F}\Big\}.$$

A small velocity of the class $\mathcal{F}$ means that it contains only gradual and slow changes in preferences, whereas a large velocity also permits abrupt changes.



(a) Mid magnitude, small velocity. (b) High magnitude, small velocity. (c) Small magnitude, high velocity.

Figure 1: Evolution of customers' preferences over time t from 0 to 10. The time required for $\big(\nu_t : t \in [0,10]\big)$ to transition from $\nu = 0$ to $\nu = 10$ illustrates the velocity of the change in preferences. A large $\mathcal{V}(\mathcal{F},T)$ allows abrupt shifts (Figure 1c), whereas a smaller value indicates that $\mathcal{F}$ contains more gradual evolution (Figures 1a and 1b).

Figure 1 illustrates how different magnitudes and velocities shape the class of admissible preferences' dynamics through representative examples that the class may contain. The magnitude and velocity of a class $\mathcal{F}$ satisfy $\mathcal{V}(\mathcal{F},T) \le \mathcal{M}(\mathcal{F},T) \le T \cdot \mathcal{V}(\mathcal{F},T)$. Thus, for a fixed horizon T, smaller velocities typically correspond to smaller magnitudes. In what follows, we present results and discussion in terms of magnitude, noting that analogous insights can be stated in terms of velocity as well.

4.2 Fundamental limit on the price of robustness

We establish a fundamental limit on the price of robustness with respect to the class of admissible preferences' dynamics $\mathcal{F}$. Our lower bound result is constructive and relies on the MNL choice model. Accordingly, and to simplify the exposition, we assume that $\mathcal{F}$ includes MNL preferences. Nonetheless, the underlying arguments can extended beyond the MNL choice model and can be used to derive similar bounds for other random-utility choice models.

Theorem 1. *Suppose $\mathcal{F}$ includes MNL preferences. Then, for $T \ge 2$, we have that:*

$$\mathcal{R}^*\big(\mathcal{F},T\big) \ge \frac{\sqrt{2}-1}{(16)^2\sqrt{2}e} T^{3/4} \mathcal{M}(\mathcal{F},T)^{1/4}.$$

Theorem 1 is obtained by constructing a deliberately challenging instance in which no assortment strategy can achieve consistently "good" performances. To develop this instance, we use an MNL choice model with attraction parameters that systematically alternate between two distinct regimes, as in Example 1. Specifically, we divide the time horizon T into carefully sized sub-segments. For each sub-segment, we randomly assign one of the two customers' preferences with equal probability. This setup creates a challenge for any policy: at least one sub-segment inevitably suffers from insufficient "forgetting," thereby pushing the regret upward.

If the magnitude of the class $\mathcal{F}$ is bounded, namely $\mathcal{M}(\mathcal{F},T) = \mathcal{O}(1)$, then Theorem 1 establishes that the regret of any assortment strategy is in the order of $\mathcal{O}(T^{3/4})$ at best. Thus, our lower bound is larger than the classical $\mathcal{O}(\sqrt{T})$ regret achieved in settings with static preferences (Agrawal et al. 2019). Note that our lower bound differs from that of order $\mathcal{O}(T^{2/3}M_T^{1/3})$ from Besbes et al. (2015) in non-stationary stochastic optimization (where M_T is a measure of the environment's variability). The key difference between their result and ours comes from their choice of measuring the environment's variability in terms of the infinite norm of the difference between consecutive cost functions, whereas we employ the KL divergence when measuring the magnitude.

Suppose instead that the magnitude of the class $\mathcal{F}$ scales as $\mathcal{M}(\mathcal{F},T) = \mathcal{O}\left(T^{\alpha}\right)$ for some $\alpha \in (0,1)$. Theorem 1 then establishes that regret is bounded below by $\mathcal{O}\left(T^{(3+\alpha)/4}\right)$. This bound characterizes the best achievable performance when balancing the remembering-forgetting trade-off. Without forgetting, regret grows linearly in T, whereas the lower bound implies an achievable regret of order $T^{3/4}$ when $\alpha = 0$. As α approaches 1, the magnitude becomes linear in T and no policy can achieve sublinear regret. In this regime, the price of robustness is linear in the horizon: even the optimal strategy cannot hedge against the volatility of preferences without incurring a worst-case opportunity cost of order T, corresponding to the cost of failing to adapt to evolving preferences.

Therefore, when the retailer imposes no restriction on the class $\mathcal{F}$ and seeks robustness against all possible dynamics, robustness carries a prohibitive price since sublinear regret is unattainable. Conversely, richer *a priori* information that tightens $\mathcal{F}$ may reduce $\mathcal{M}\left(\mathcal{F},T\right)$ and lower that price.

4.3 Robustness through pre-planned restarts

Next, we consider a restart-based strategy that applies a policy designed for static preferences over successive epochs and re-initializes it between epochs. Under this approach, the retailer's decision reduces to selecting an appropriate restart timing, which determines how much past data the policy retains. Theorem 1 provides guidance for selecting this timing so as to match the lower bound limit on performance, thereby optimally balancing the remembering-forgetting trade-off.

Algorithm 1 Restart-based strategy $\pi\left(\Delta,\mathcal{A}\right)$

```
Input: A batch-size Δ and a policy A for the static setting
while 1 ≤ j ≤ ⌈T/Δ⌉ do
    Run A on consumer t = (j − 1) Δ + 1 to t = min{j Δ, T}          (restart)
    j = j + 1
```

The restart-based strategy, depicted in Algorithm 1, applies a dynamic assortment planning policy $\mathcal{A}$ designed for static preferences over consecutive blocks of Δ periods and restarts it at the beginning of each block. This approach allows the retailer to leverage the rich and growing literature on dynamic assortment planning under static preferences (Sauré and Zeevi 2013; Agrawal et al. 2019; Li et al. 2025). At each restart, all past observations are discarded and learning is re-initialized using only data collected within the current block, ensuring that assortment decisions are based on the most recent estimates of customers' preferences.

We evaluate the performance of Algorithm 1 in terms of the regret incurred by $\mathcal{A}$ under worst-case static preferences. To this end, we define the set of admissible static preferences as:

$$\mathcal{F}_S := \left\{ F^{(\mathbb{N})} \in \mathcal{F} : \, F^t = F^{t-1}, \, t > 1 \right\},$$

and denote by $\mathcal{R}^{\mathcal{A}}(\mathcal{F}_S, \Delta)$ the corresponding regret when serving Δ customers. Next, we derive an upper bound on the regret of the restart-based strategy for any restart timing Δ, expressed in terms of both the regret $\mathcal{R}^{\mathcal{A}}(\mathcal{F}_S, \Delta)$ for static preferences and the magnitude $\mathcal{M}(\mathcal{F}, T)$ of the class $\mathcal{F}$.

Theorem 2. *For $\mathcal{A} \in \mathcal{P}$ and $\Delta \leq T$, let $\pi \equiv \pi(\Delta, \mathcal{A})$ be the policy defined in Algorithm 1 and $\mathbf{w} \equiv \left(w_i \, : \, i \in \mathcal{N}\right)$. Then, for $T \geq 2$,*

$$\mathcal{R}^{\pi}\left(\mathcal{F}, T\right) \leq \lceil T/\Delta \rceil \cdot \mathcal{R}^{\mathcal{A}}\left(\mathcal{F}_S, \Delta\right) + \left(N+1\right)\|\mathbf{w}\|_1 \cdot \sqrt{T\Delta/2} \cdot \mathcal{M}(\mathcal{F}, T)^{1/2}.$$

The upper bound on the regret in Theorem 2 is obtained by introducing an intermediate benchmark, which we term a *semi-oracle*. The semi-oracle has full knowledge of customers' preferences within each Δ-length subsegment but is restricted to choosing a single assortment for all customers in that subsegment. Using this benchmark, we decompose the regret of the restart-based strategy into *(i)* the regret relative to the semi-oracle, capturing the cost of learning within each restart block, and *(ii)* the regret of the semi-oracle relative to the clairvoyant retailer, which quantifies the loss from committing to a fixed assortment over a block when preferences evolve within it.

The bound in Theorem 2 decomposes into two terms that capture distinct sources of regret arising from robustness to evolving preferences in $\mathcal{F}$. The first term scales with $\mathcal{R}^{\mathcal{A}}\left(\mathcal{F}_S, \Delta\right)$ and depends solely on the performance of $\mathcal{A}$ under static preferences, through comparison with the semi-oracle. Prior work provide worst-case regret guarantees in the static setting (Agrawal et al. 2019; Li et al. 2025), so any improvement in those results immediately tighten our bound. In particular, this static component can be as small as $\mathcal{O}(\log T)$ under milder robustness requirements, but grows to $\mathcal{O}(\sqrt{T})$ when the retailer hedges against arbitrarily hard static instances. The second term scales with the magnitude $\mathcal{M}\left(\mathcal{F}, T\right)$ and reflects the retailer's risk profile. With weaker *a priori* restrictions on $\mathcal{F}$, the magnitude $\mathcal{M}(\mathcal{F}, T)$ can grow arbitrarily large, enlarging the advantage of the clairvoyant retailer and allowing preferences to vary substantially within each Δ-length segment.

These two components interact through the restart timing Δ, which the retailer controls directly. Choosing Δ balances two opposing effects. A smaller Δ allows the policy to react more quickly to evolving preferences, but induces more frequent restarts, limits learning within each segment, and may result in excessive forgetting. Conversely, a larger Δ improves estimation accuracy within each segment, but delays adaptation to changes in preferences and may lead to excessive remembering. Accordingly, Δ must be small enough to ensure timely adaptation, yet large enough to support meaningful learning within each segment. Balancing the remembering-forgetting trade-off therefore reduces to selecting an appropriate restart timing Δ.

Corollary 1. *For $\mathcal{A} \in \mathcal{P}$ and $\Delta \equiv \lceil T^{1/2}\mathcal{M}(\mathcal{F},T)^{-1/2}\rceil$, let $\pi \equiv \pi(\Delta, \mathcal{A})$ be the policy defined in Algorithm 1 and $\mathbf{w} \equiv \big(w_i : i \in \mathcal{N}\big)$. Then, for $T \geq 2$,*

$$\mathcal{R}^{\pi}\big(\mathcal{F}, T\big) \leq 2T^{1/2}\mathcal{M}(\mathcal{F},T)^{1/2} \cdot \mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_S, \Delta\big) + (N+1)\|\mathbf{w}\|_1 \cdot T^{3/4} \cdot \mathcal{M}(\mathcal{F},T)^{1/4}.$$

A strategic choice emerges when Δ is set to $\mathcal{O}(T^{1/2}\mathcal{M}(\mathcal{F},T)^{-1/2})$, which leads to the upper bound from Corollary 1. With this choice of Δ, and provided that $\mathcal{A}$ incurs a regret of $\mathcal{O}\big(\sqrt{\Delta}\big)$ over each sub-segment, the restart-based strategy achieves near-optimal performances, "almost" matching the lower bound on regret from Theorem 1. As both the horizon T and the magnitude $\mathcal{M}(\mathcal{F}, T)$ increase, the volatility of the class $\mathcal{F}$ also increases. In response, reducing the size Δ ensures the algorithm adapts frequently enough to manage this increased volatility.

4.4 Operating with evolving customers' preferences

The lower bound on achievable performance in Theorem 1 highlights the intrinsic difficulty of operating with customers whose preferences evolve. To make this effect more concrete, we construct a class of admissible preferences' dynamics with varying magnitude and illustrate numerically how the retailer's risk profile, through its magnitude, directly governs the price of robustness.

Example 2. We consider a sequence of settings, in which the horizon T ranges from 1 to 10000. For $\alpha \in \{0, 0.5, 0.75\}$, we define $M_T = \frac{1}{8}T^{\alpha}$ and we use the class $\mathcal{F}_{\text{MNL}}(M_T)$ of evolving preferences as in Example 1. Next, we set $\Delta = \lceil T^{1/2} M_T^{-1/2}\rceil$ as an input of Algorithm 1. Also, we use the policy $\mathcal{A}$ by Agrawal et al. (2019) as a subroutine within our restart-based strategy. ■

Figure 2 illustrates the challenge faced by the retailer when preferences evolve. As α increases, the magnitude M_T grows, allowing preferences within $\mathcal{F}_{\text{MNL}}(M_T)$ to vary more frequently (recall

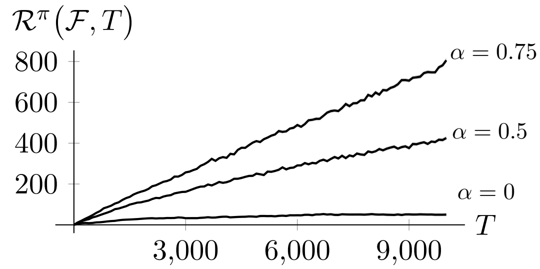


Figure 2: Regret of the restart-based strategy $\pi(\Delta, \mathcal{A})$ described in Algorithm 1 applied to the sequence of settings from Example 2 with $M_T = T^{\alpha}$, for $\alpha \in \{0, 0.5, 0.75\}$, where the horizon ranges from $T = 1$ to $T = 10000$. We compute the average regret (in black) and the 95% confidence interval for the mean (imperceptible) over 500 instances.

Example 1). Maintaining robustness under this heightened volatility therefore requires the retailer to restart more frequently. Moreover, Corollary 1 provides a clear economic implication.

Robustness against a broad class of dynamics requires hedging against highly volatile preferences, which widens the information gap between the retailer and an adversary that selects $F^{(\mathbb{N})}$. As the class of admissible preferences' dynamics $\mathcal{F}$ expands, the adversary's choice set enlarges, amplifying this gap, increasing M_T, and, in turn, the retailer's regret. Pre-planned restarts can control this regret, but they do not exploit any structure in the preferences' dynamics beyond the magnitude of the class $\mathcal{F}$. This observation raises a natural question: can the retailer trade some robustness by restricting the class $\mathcal{F}$ for exploitable structure in preferences' dynamics in order to better balance the remembering-forgetting trade-off, rather than relying solely on pre-planned restarts?

5 How risk profiles shape robust planning

Motivated by shocks such as pandemics or financial crises, and by short product lifecycles in fast fashion where slow evolution of preferences has limited effect, we focus on settings in which the retailer seeks robustness only to abrupt changes in preferences. This more constrained risk profile leads to a distinct robust planning regime, where the key challenge shifts from carefully scheduling pre-planned restarts to detecting abrupt changes and responding promptly. Although the magnitude still provides a broad guide for robust planning, the velocity becomes critical. Two risk profiles with identical magnitude can require different strategies and result in different regret outcomes.

5.1 Robustness to high-velocity preferences' dynamics

We consider abrupt changes, focusing on risk profiles in which customers' preferences remain static for a period and then shift at a single point. Formally, we define a class $\mathcal{F}_A \subseteq \mathcal{F}$ of admissible preferences' dynamics that are constant except for a single change point $\tau \in \mathbb{N}$:

$$\mathcal{F}_A := \left\{ F^{(\mathbb{N})} \in \mathcal{F} \,:\, F^t = F^{t-1},\, \forall\, t \neq \tau > 1,\, \tau \in \mathbb{N} \right\}.$$

From this definition, it follows that both the timing τ and the magnitude of the change are unknown by the retailer. This class is characterized by high velocity, in the sense that $\mathcal{M}(\mathcal{F}_A, T) = \mathcal{V}(\mathcal{F}_A, T)$.

We refer to F^1 and F^τ as the *pre-* and *post-change preferences*, respectively. Also, we make two assumptions regarding this setting. First, we assume that the retailer knows the initial preferences; this assumption is mild and helps us isolate the challenge of adapting to a change from learning the initial preferences. Second, we assume that post-change preferences within class $\mathcal{F}_A$ are asymptotically "well separated" in the sense that the *minimum optimality gap* as defined by:

$$\gamma \equiv \gamma(\mathcal{F}_A) := \inf\Big\{ r\big(S^*(F^t), F^t\big) - r(S, F^t) \,:\, F^{(\mathbb{N})} \in \mathcal{F}_A \,,\, t \in \mathbb{N} \,,\, S \in \mathcal{S} \,,\, S \neq S^*(F^t) \Big\},$$

is strictly positive, i.e., $\gamma > 0$. This assumption is technical and is used to derive a lower bound on the achievable performance of any robust planning strategy. We impose no further restriction on γ, so it may either depend on the horizon T or be independent of it.

Since initial preferences are known, when needed, we denote by $\mathcal{F}_A(F^1)$ the subset of preferences from $\mathcal{F}_A$ in which the initial ones are given by F^1. Accordingly, we define the worst-case performance across all possible pre-change preferences as:

$$\widetilde{\mathcal{R}}(\mathcal{F}, T) \;\equiv\; \sup\left\{ \mathcal{R}^*(\mathcal{F}_A(F^1), T) :\, F^1 \in \mathcal{F} \right\},$$

and assess the performance of a policy π via $\mathcal{R}^\pi(\mathcal{F}_A(F^1), T)$. We assume that any shift results in a change in the single-sale optimal assortment, i.e., $S^*(F^1)$ and $S^*(F^\tau)$ differs by at least one product.

5.2 Large abrupt changes and active monitoring

We begin by examining scenarios in which changes in preferences might go *undetected* by the retailer. Such shifts occur when preferences change without affecting the products in the pre-change optimal assortment $S^*(F^1)$. In these cases, a retailer that continues to offer $S^*(F^1)$ would observe no change in the customers' purchasing behavior. For example, in fashion, an assortment consisting only of shirts may fail to reveal shifts in preferences for pants or shoes, even if those products become more profitable after the change. We refer to these shifts as *passively undetectable.*

We model the retailer as expecting the size of any change to lie within a known range. An upper bound on the size of the change ensures that the magnitude of the class does not grow with the horizon, i.e., $\mathcal{M}(\mathcal{F}_A, T) \leq M$ for some constant $M > 0$. This reflects the idea that longer horizons do not make physical shocks, such as pandemics or financial crises, intrinsically larger. A lower bound on the size of the change reflects operational considerations: the retailer may seek robustness only to sufficiently large changes, since reacting to minor shifts can be impractical.

Formally, we define the following class of preferences' dynamics:

$$\tilde{\mathcal{F}}_U := \left\{F^{(\mathbb{N})} \in \mathcal{F}_A \,:\, \mathcal{K}^\tau(S^*_{\tau-1}) = 0, \, \max\left\{\mathcal{K}^\tau(S) : S \in \mathcal{S}\right\} \in (\kappa, \phi)\right\},$$

for $\kappa \in (0,1)$ and $\phi > \kappa$ defining a range for the magnitude $\mathcal{M}(\tilde{\mathcal{F}}_U, T)$. The first condition ensures that changes are passively undetectable, while the second, a *separability* condition, guarantees the existence of an assortment where preferences are sufficiently distinct to be observed and acted upon.

5.2.1 Small changes are the real threat

We now establish a lower bound on the performance of any assortment strategy when the change is both passively undetectable and *separable*. Our result parallels that of Besbes and Zeevi (2011) in dynamic pricing and draws on probabilistic techniques from Tsybakov (2003). However, in our setting, the combinatorial nature of assortment planning requires different technical arguments.

Proposition 1. *There exists some finite constant $C \equiv C(\gamma, \phi) > 0$, such that, for $T \geq 2$:*

$$\widetilde{\mathcal{R}}(\tilde{\mathcal{F}}_U, T) \geq C\sqrt{T}.$$

The bound in Proposition 1 is derived constructively by designing a worst-case instance and highlights the remembering-forgetting trade-off through a game-theoretic lens. In this game, the retailer first commits to a strategy, and an adversarial environment then selects both the timing and the extent of the change in response to that strategy. If the strategy forgets too aggressively, then the change is placed near the final period, since the retailer never fully trusts its sales data and must repeatedly relearn preferences. Conversely, if the retailer relies too much on past sales data over a segment, then the environment positions the change in that segment to exploit insufficient forgetting.

Proposition 1 reveals an unusual dependence on $\mathcal{M}(\mathcal{F}_U, T)$ through ϕ: the lower bound increases when the magnitude decreases. Under the more adversarial environment from Theorem 1, tightening $\mathcal{F}$ reduces $\mathcal{M}(\mathcal{F}, T)$ and simplifies the problem. Here, the opposite occurs. Restricting the dynamics to limit volatility may still permit arbitrarily small shifts. If κ is very small, the constructed worst-case shift is subtle enough to evade detection yet persistent enough to steadily erode revenue.

5.2.2 From pre-planned restarts to active monitoring

The nature of the retailer's risk profile, together with the separability condition, enables a shift from pre-planned restart strategies to active monitoring approaches. We introduce an active monitoring strategy (Algorithm 2) that balances the remembering-forgetting trade-off by retaining historical information as long as no change is detected, and discarding it only when necessary.

The policy alternates between *sentinel* and *steady* cycles of lengths $\Delta_e = \mathcal{O}(\log T)$ and $\Delta_o = \mathcal{O}(\sqrt{T})$, respectively. During each sentinel cycle, the retailer offers assortments from $\mathcal{E} \subseteq \mathcal{S}$, which are designed to reveal a change through statistical testing. If no change is detected, then the retailer enters the subsequent steady cycle and continues to offer the pre-change optimal assortment, thereby preserving past information. Otherwise, the retailer discards pre-change information and runs a dynamic assortment planning policy $\mathcal{A}$ for the remainder of the horizon to learn the new preferences, so the strategy restarts only when a change is confirmed.

Algorithm 2 Active monitoring strategy $\pi\left(\kappa, F^1, \mathcal{E}, \mathcal{A}\right)$

Input: A constant $\kappa > 0$, a distribution F^1, a set of test assortments $\mathcal{E}$, and a policy $\mathcal{A}$
Initialize: Set $detect = False$, $t = 0$, $\Delta_o := \sqrt{T}/\kappa^2$, $\Delta_e := 4(\log T)/\kappa^2$
while $detect = False$ and $t \leq T$ **do**
  Offer $S^u = S^*(F^1)$ for $u = t+1, \ldots, t+\Delta_o$ (steady cycle)
  Offer each assortment $S \in \mathcal{E}$ to Δ_e customers (sentinel cycle)
  if $\left|\sum_{u=t+1}^{t+\Delta} \mathbf{1}\{i^u = i\} - p_i(S, F^1)\right| > \Delta\,\kappa/2$ for some $i \in S \cup \{0\}$, and $S \in \mathcal{E}$ **then**
    $detect = True$ (change detected)
  $t = t + \Delta_o + |\mathcal{E}|\Delta_e$
Run $\mathcal{A}$ on customers $t+1$ to T (post-change policy)

This assortment strategy differs fundamentally from Algorithm 1, as it actively seeks to detect whether a change in preferences has occurred. In particular, it employs a statistical test to determine whether the deviation between the empirical purchasing probabilities and those expected from pre-change preferences is "abnormally" large. Proposition 2 bellow establishes an upper bound on the regret of this strategy. There, we assume that $\mathcal{E}$ includes the assortment satisfying the separability condition in $\tilde{\mathcal{F}}_U$ (we provide further details on the selection of $\mathcal{E}$ in our Appendix A.1).

Proposition 2. *For $\kappa > 0$, F^1 such that $F^{(\mathbb{N})} \in \tilde{\mathcal{F}}_U$ and $\mathcal{A} \in \mathcal{P}$, let $\pi \equiv \pi\left(\kappa, F^1, \mathcal{E}, \mathcal{A}\right)$ be the policy defined in Algorithm 2. Then, there exists finite constants $C_1 \equiv C_1(K, \kappa, \mathcal{E}) > 0$, $C_2 \equiv C_2(\kappa, \mathcal{E}) > 0$ and $t \equiv t(\kappa, K)$, such that, for $T \geq t$:*

$$\mathcal{R}^\pi(\tilde{\mathcal{F}}_U(F^1), T) \leq C_1 + C_2 \log T + 4\|\mathbf{w}\|_1 |\mathcal{E}| \sqrt{T} \log T + \mathcal{R}^{\mathcal{A}}\big(\tilde{\mathcal{F}}_b(F^1), T\big),$$

where $\tilde{\mathcal{F}}_b(F^1) := \left\{\tilde{F}^{(\mathbb{N})} \in \mathcal{F}_S \,:\, \tilde{F}^1 = G^\tau \,,\, G^{(\mathbb{N})} \in \tilde{\mathcal{F}}_U(F^1)\right\}$.

The regret upper bound in Proposition 2 decomposes into three components, each capturing a different source of regret for our active monitoring strategy. First, the $\mathcal{O}(\log T)$ term reflects the detection delay of the statistical test, namely the time needed to accumulate enough evidence that preferences have shifted. Second, the $\mathcal{O}(\sqrt{T}\log T)$ term captures the cost of maintaining exploration over assortments in $\mathcal{E}$, which becomes unnecessary when the change occurs late in the horizon. Finally, the third term accounts for the regret incurred while learning the preferences after the change occurs to find the post-change optimal assortment.

The decomposition in Proposition 2 illustrates that the source of difficulty differs from the general setting of Theorem 1. In the latter, broader classes of admissible preferences' dynamics lead to a larger magnitude of the class $\mathcal{F}$, which in turn amplifies the challenge for the retailer.

In the more constrained setting that motivates monitoring-based policies (Proposition 1), the magnitude is bounded. The difficulty instead arises from changes that are small enough to be hard to detect, yet large enough to significantly erode revenue. Thus, the operational challenge shifts from coping with highly volatile preferences to carefully detecting subtle but impactful shifts.

Proposition 2 highlights a trade-off between testing for a change and relearning preferences to determine the appropriate planning regime. A restart-based strategy with an appropriate choice of restart timing can achieve regret of the same order, so the value of change detection depends on whether testing is cheaper than restarting. In particular, a change-detection approach is attractive when a shift can be revealed through assortments that are less costly than relearning preferences from scratch. This attractiveness depends on the retailer's risk profile through the class $\tilde{\mathcal{F}}_U$.

Indeed, the potential gain from exploiting the abrupt-change structure is reflected in the constant multiplying the third term of order $\mathcal{O}(\sqrt{T})$ through $|\mathcal{E}|$. This dependence is linear in the size of the set of test assortments and captures how testing cost depends on the class $\tilde{\mathcal{F}}_U$, namely, how many assortments are required to reveal a change. If correlations across products allow shifts to be revealed through a small collection of test assortments, then change detection can materially improve over restart-based approaches. By contrast, in settings in which preferences are governed by MNL choice models, detecting a change may require exploratory effort of the same order as periodically relearning preferences, consistent with the exploration requirements in Sauré and Zeevi (2013).

5.3 Detectability: similar risk profiles but different planning strategies

We continue to consider robustness to a single large abrupt shift in customers' preferences, but we now highlight a source of heterogeneity across retailers. Even under similar risk profiles, the impact of a change depends on whether it is detectable from the assortment the retailer routinely offers. We therefore focus on settings where the retailer expects its pre-change optimal assortment to be informative enough to reveal the shift through observed sales.

This situation is of particular interest when the retailer carries a diverse set of products, so that the pre-change optimal assortment remains representative of the broader product set. To formalize this idea, we define the class of preferences' dynamics with *passively detectable* changes as follows:

$$\mathcal{F}_D := \left\{F^{(\mathbb{N})} \in \mathcal{F}_A \,:\, \mathcal{K}^{\tau}(S^*_{\tau-1}) > \varepsilon\right\},$$

where $\varepsilon \in (0,1)$ provides a lower bound on the magnitude of the class $\mathcal{F}_D$. Changes in preferences within $\mathcal{F}_D$ are large enough on products in the pre-change optimal assortment to be detected *passively* by continuing to offer that assortment and monitoring purchasing data. In particular, the risk profiles induced by the class $\mathcal{F}_D$ above and the class $\tilde{\mathcal{F}}_U$ from Section 5.2 are similar when $\varepsilon = \kappa$.

5.3.1 Fundamental performance limit

We now derive a fundamental performance bound for any assortment strategy when the change can be detected passively. Our result parallels that of Besbes and Zeevi (2011) in the context of dynamic pricing with abruptly changing demand, though key differences arise due to the non-convexity of the assortment planning problem, which necessitates different technical arguments.

Proposition 3. *There exists constants $C \equiv C(\gamma, \varepsilon) > 0$ and $t \equiv t(\gamma, \varepsilon) > 0$, such that, for $T \geq t$:*

$$\widetilde{\mathcal{R}}^*(\mathcal{F}_D, T) \geq C \log T.$$

The result follows from a constructive argument that selects an adversarial change point τ for any given assortment strategy. Specifically, an adversary environment can choose τ so that the strategy fails to offer the post-change optimal assortment within a $\mathcal{O}(\log T)$-length window around the change point τ, which in turn forces a regret of the strategy to be of order $\mathcal{O}(\log T)$.

5.3.2 From active to passive monitoring

We describe a passive monitoring approach that exploits the detectability of changes in preferences (see Algorithm 3). In this conceptual strategy, the retailer operates in cycles of length $\Delta = \mathcal{O}(\log T)$, during which deviations in purchasing frequencies within the pre-change optimal assortment are monitored. If no change is detected, then the pre-change optimal assortment is maintained; otherwise, the retailer updates its policy to learn the new preferences. Unlike Algorithm 2, this approach does not require testing alternative assortments.

The next result establishes an upper bound on the regret of the passive monitoring strategy. In particular, it shows that the strategy attains the best achievable performance from Proposition 3, up to the additional regret incurred while learning the new preferences.

Algorithm 3 Passive monitoring strategy $\pi\left(\varepsilon, F^1, \mathcal{A}\right)$

Input: A constant $\varepsilon > 0$, a distribution F^1, and a policy $\mathcal{A}$ for the static setting
Initialize: Set $detect = False$, $t = 0$, $\Delta = 4(\log T)/\varepsilon^2$
while $detect = False$ and $t < T$ **do**
  Offer $S^u = S^*\left(F^1\right)$ for $u = t+1, \ldots, t+\Delta$ (passive monitoring)
  if $\left|\sum_{u=t+1}^{t+\Delta} \mathbf{1}\{i^u = i\} - p_i(S^*(F^1), F^1)\right| > \Delta\,\varepsilon/2$ for some $i \in S^*(F^1) \cup \{0\}$ **then**
    $detect = True$ (change detected)
  Set $t = t + \Delta$
Run $\mathcal{A}$ on customers $t+1$ to T (post-change policy)

Proposition 4. *For $\varepsilon > 0$, F^1 such that $F^{(\mathbb{N})} \in \mathcal{F}_A$ and $\mathcal{A} \in \mathcal{P}$, let $\pi \equiv \pi\left(\varepsilon, F^1, \mathcal{A}\right)$ be the policy defined in Algorithm 3. Then, there exists finite constants $C_1 \equiv C_1(\varepsilon) > 0$, $C_2 \equiv C_2(\varepsilon) > 0$ and $t \equiv t(\varepsilon, K) > 0$, such that, for $T \geq t$:*

$$\mathcal{R}^\pi(\mathcal{F}_D(F^1), T) \leq C_1 + C_2 \log(T) + \mathcal{R}^{\mathcal{A}}\left(\mathcal{F}_b(F^1), T\right),$$

where $\mathcal{F}_b(F^1) := \left\{\tilde{F}^{(\mathbb{N})} \in \mathcal{F}_S \,:\, \tilde{F}^1 = G^\tau\,,\, G^{(\mathbb{N})} \in \mathcal{F}_D(F^1)\right\}$.

The last term in the bound of Proposition 4 captures the regret incurred by the policy $\mathcal{A}$ when learning the post-change preferences. The first term reflects the regret due to detection delays or errors from the statistical test. When $\mathcal{A}$ achieves $\mathcal{O}(\log T)$ regret, as in the well-separated setting of Sauré and Zeevi (2013), Proposition 4 implies that the overall regret remains of order $\mathcal{O}(\log T)$. In this case, evolving preferences affect performance only through the constant multiplying $\log T$.

This observation highlights that retailers facing similar risk profiles need not rely on the same robust planning regimes. Under large abrupt changes, a retailer whose pre-change optimal assortment provides sufficient coverage can treat change detection as a passive monitoring task: it can keep offering its pre-change optimal assortment, track purchases, and restart learning only when the data reveal a shift. The practical implication is that the main lever becomes the quality of the statistical test rather than additional algorithmic sophistication in the assortment policy. By contrast, when coverage is limited and changes are not passively detectable, the retailer must rely on active testing or periodic restarts, and a similar risk profile translates into a markedly higher opportunity cost.

5.4 Trading robustness for structural information

The preceding sections illustrate how structural information about changes in customers' preferences, articulated through a risk profile, can materially improve assortment planning. By shrinking the set of admissible dynamics against which the retailer seeks robustness, the regret's dependence on the horizon improves from $\mathcal{O}(T^{3/4})$ in the generic case of Section 4 to at most $\mathcal{O}(T^{1/2})$. Beyond this improved rate, our results show that additional structure can also justify a shift in the planning regime from pre-planned restarts to proactive monitoring. By tracking deviations and updating assortments only when a change is detected, the retailer avoids forgetting when it is unnecessary. Such monitoring is particularly valuable when testing is less costly than relearning preferences.

This advantage is especially pronounced when the change is detectable from the pre-change optimal assortment. In that case, the retailer can continue offering $S^*(F^1)$ while monitoring sales data, so detection requires no operational disruption. Once the data indicate a shift, the retailer can adapt promptly, avoiding the prolonged mismatch that arises from offering an assortment based on outdated preferences. As a result, the additional opportunity cost is only of order $\mathcal{O}(\log T)$, on top of the cost of learning post-change preferences, which is substantially smaller than the $\mathcal{O}(\sqrt{T})$ regret associated with passively undetectable changes.

Taken together, our results indicate that balancing the remembering-forgetting trade-off is not primarily an algorithmic question, but a strategic one. Performance depends, first, on how the retailer articulates its risk profile given the products it carries and the assortments it can offer, and second, on how it implements that risk profile through its assortment stategy design.

6 Case study: planning under evolving customers' preferences

We use clickstream data from a Chilean retailer to construct preferences' dynamics that reflect two market scenarios. We use these two scenarios to illustrate how the risk profiles and planning choices developed throughout the study apply. Specifically, we study preferences' evolution under (i) seasonal diffusion in fashion, and (ii) an abrupt shift triggered by a sudden shock. The first scenario quantifies the cost of inaction and shows the value of adapting assortments as preferences evolve, relative to applying a policy designed for static preferences. The second scenario illustrates how a tighter risk profile can guide more targeted restarts and improve the retailers' performance.

6.1 Implementation details

In our analysis, we use a clickstream dataset from a Chilean retailer, comprising approximately 94,000 customer interactions. In each interaction, customers are presented with an assortment of $K = 4$ products from a portfolio of 19 items. Customers are segmented into 42 demographic profiles defined by gender, age group, and geographic region. A comprehensive description of the dataset is provided in Bernstein et al. (2019), who originally used it in the context of dynamic assortment planning with personalization. Further details on our implementation can be found in Appendix A.2.

Customers' preferences. We calibrate preferences using an MNL choice model, leveraging data from specific sub-groups of the full dataset (e.g., customers from a particular region) to construct our scenarios. The sub-groups used for calibration are specified at the beginning of each scenario. The attraction parameter for the no-click option is set to 1, and each product $i \in \mathcal{N}$ yields a profit of $w_i = 1$. We estimate the attraction parameters using the estimator proposed by Bernstein et al. (2019) in a similar context. Specifically, for each product $i \in \mathcal{N}$, it is defined as:

$$\hat{\nu}_i = \frac{\sum_t \mathbf{1}\big(i_t = i \text{ and } i \in S^t\big)}{\sum_t \mathbf{1}\big(i_t = 0 \text{ and } i \in S^t\big)},$$

where S^t denotes the assortment shown to customer t.

Experimental setup. We normalize the attraction parameters $\hat{\nu}$ by their maximum values, so that $\max\{\hat{\nu}_i : i \in \mathcal{N}\} = 1$. This scaling reduces the proportion of no-clicks during both the learning and the change detection procedures, thereby significantly accelerating the convergence of our procedures while preserving the relative ranking of the parameters. Since the profit for each product is identical, the optimal assortments remain unchanged after scaling the attraction parameters.

6.2 The cost of *not* planning for evolving customers' preferences

Our goal is to numerically illustrate the cost of not planning for evolving customers' preferences. We use the entire dataset to calibrate a scenario that captures the seasonal evolution of customers' preferences in the footwear market. Seasonality is a key factor in retail and has been studied by Caro and Gallien (2007) in the context of dynamic assortment planning. We first calibrate some initial preferences $\hat{\nu}^1$ with the entire data. Note that, the corresponding initial optimal assortment contains three long boots. This estimate reflects preferences of customers for long, insulated boots (left panel of Figure 3) ideal for winter conditions.

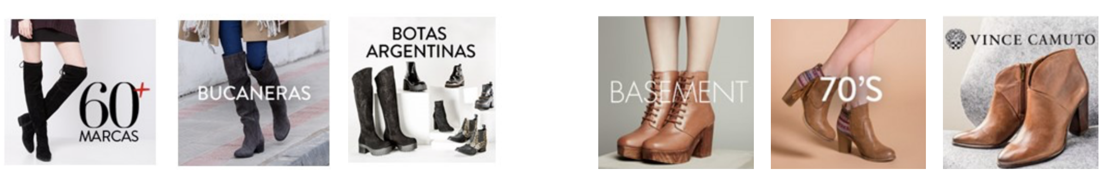


Figure 3: Left panel shows three high boots (darker tones) from the initial optimal assortment. Right panel shows three shorter shoes that gradually replace them over time.

To construct our scenario, we postulate that, as the season moves toward summer, customers' preferences gradually move toward shorter and lighter shoes that are more comfortable in warmer weather (right panel of Figure 3). While such evolution may be obvious in this setting, it need not be obvious to a retailer in general, which motivates our approach of not imposing assumptions on the preferences' dynamics such as explicit seasonality. We model the evolution as a gradual shift from the initial preferences defined earlier toward a new set of preferences with attraction parameters $\hat{\nu}^T$.

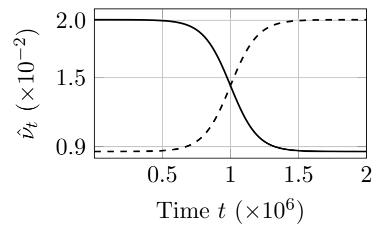


Figure 4: Attraction parameters evolution for Botas Argentinas (solid line) and Vince Camuto (dashed line). The evolution is governed by $s_t = (1 + \exp(-20\frac{t-100}{T-100} + 10))^{-1}$, with $T = 2 \times 10^6$; see Figure 3 for the product images.

To derive $\hat{\nu}^T$, we "swap" the attraction parameters for products from the left panel of Figure 3 with those from the right one. Consequently, the optimal assortment at the last time period contains lighter shoes. The gradual "swap" in preferences throughout the horizon is illustrated in Figure 4, showing the evolution of $\hat{\nu}_t$ for both *Botas Argentinas* and *Vince Camuto*. The transition follows a sigmoidal curve, beginning with a period of stability before smoothly shifting to a new regime over $T = 2 \times 10^6$ visits of customers to the retailer.

We compare two approaches for handling seasonality: the assortment strategy from Agrawal et al. (2019) designed for static preferences (denoted by $\mathcal{A}$) and the restart-based strategy π from Algorithm 1. Policy π incorporates $\mathcal{A}$ as a subroutine and determines Δ based on Corollary 1 with the appropriate magnitude. The regret for both approaches is presented in Figure 5. Moreover, we consider a sequence of settings in which the horizon T varies from $T = 1$ to $T = 2 \times 10^6$.

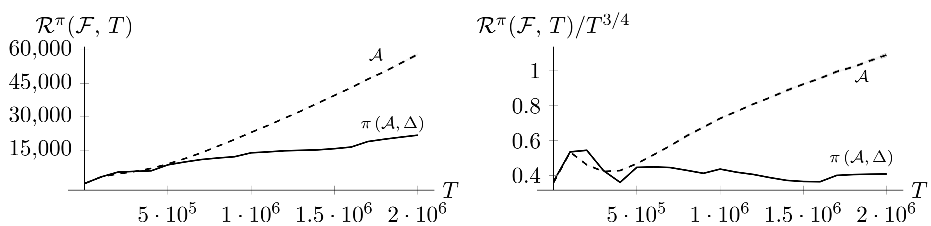


Figure 5: Regret incurred by two policies under customers' preferences that evolve from winter to summer preferences as a function of the horizon T. The *dashed* line corresponds to the policy $\mathcal{A}$ from Agrawal et al. (2019), whereas the *solid* line corresponds to the restart-based strategy $\pi(\mathcal{A}, \Delta)$ from Algorithm 1. We compute the average regret (in black), and 95% confidence interval for the mean (imperceptible) over 100 instances.

Discussion. The policy $\mathcal{A}$ performs well in the short term, matching the performance of π. Indeed, both strategies perform similarly when the horizon is low (fewer than 5×10^5), but at 2×10^6 customers, the opportunity-cost gap between them widens by a factor of three. Our results show that relying on the assortment strategy $\mathcal{A}$ becomes increasingly costly as customers' preferences evolve: its regret grows linearly, leading to substantial revenue shortfalls. In contrast, the robust planning policy π achieves sublinear regret of order $\mathcal{O}(T^{3/4})$, aligning with our theoretical predictions. This finding highlights a risk for retailers: failing to adapt assortments in response to evolving preferences can result in significant long-term missed profit. By contrast, retailers that adjust their offerings (for instance through pre-planned restarts) mitigate revenue erosion.

Beyond the numerical gains, this scenario illustrates the *central message* of the study. When the risk of evolving preferences is ignored, assortments become progressively misaligned with customers' preferences. By contrast, articulating a suitable risk profile allows the retailer to anticipate such dynamics and adapt its assortment in a controlled and predictable way. In this sense, the cost to the retailer comes not only from the change itself, but also from the absence of planning for it.

6.3 Tightening the risk profile

We consider a setting in which the retailer seeks robustness against an abrupt change in customers' preferences. Our goal is to illustrate how tightening the risk profile and exploiting passive detectability can help manage this risk. We also provide intuition on the cost of overconservatism, namely planning for dynamics that are more adverse than those the retailer actually encounters.

To this end, we construct a scenario in which preferences shift abruptly in the aftermath of a pandemic. We calibrate pre-change preferences using data from the 30-39 age range, represented by the attraction parameters $\hat{\nu}^1$ in Figure 6. When the pandemic strikes, a broader cross-section of consumers transitions from in-store to online shopping from one day to the next due to sanitary restrictions. This shock induces new attraction parameters $\hat{\nu}^\tau$, calibrated using the full dataset.

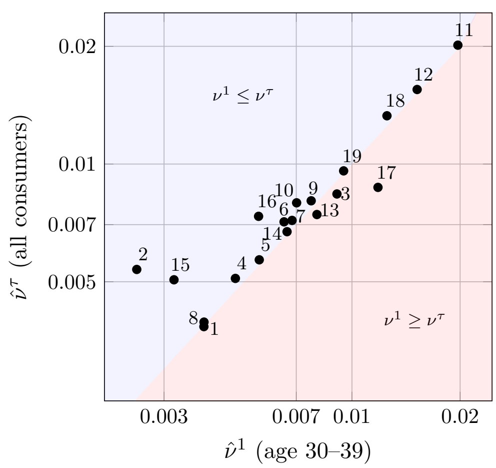
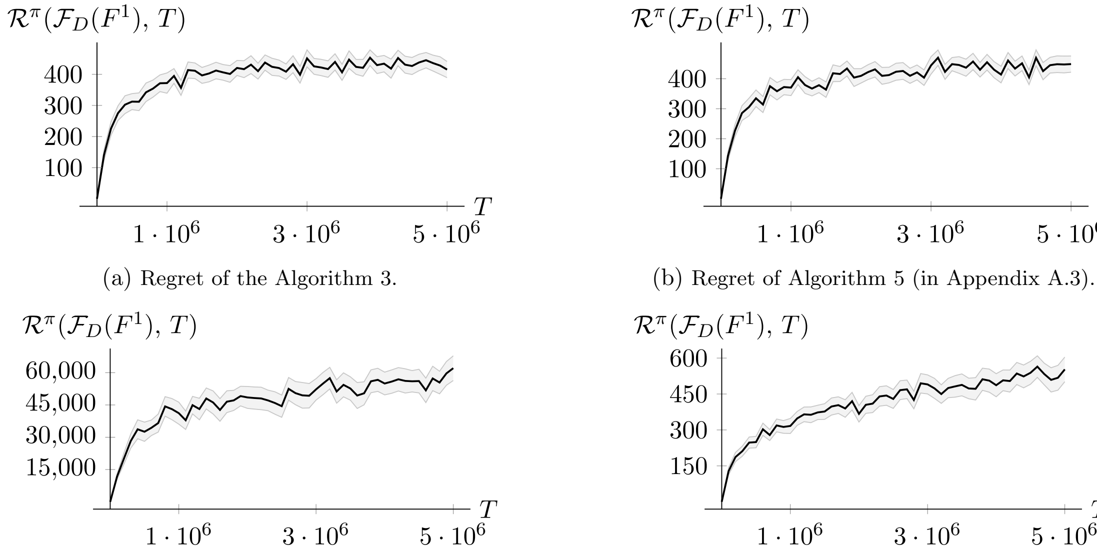


Figure 6: Estimated attraction parameters $\hat{\nu}^1$ (age 30–39) against $\hat{\nu}^\tau$ (all customers).

The optimal assortment before the change occurs is $S^*(\hat{\nu}^1) = \{11, 12, 17, 18\}$, which then shifts to $S^*(\hat{\nu}^\tau) = \{11, 12, 18, 19\}$. Additionally, we consider a sequence of settings in which the horizon ranges from $T = 1$ to $T = 5 \times 10^6$, with 100 independent replications of each setting. For each simulation, the change point τ is drawn uniformly at random over the period $[T]$. Consequently, for $t < \tau$, the attraction parameters are given by $\nu^t = \hat{\nu}^1$, and after τ, they switch to $\nu^t = \hat{\nu}^\tau$.



(a) Regret of the Algorithm 3.

(b) Regret of Algorithm 5 (in Appendix A.3).

(c) Regret of Algorithm 2.

(d) Regret of Algorithm 4 (in Appendix A.3).

Figure 7: Regret for assortment strategies when preferences change abruptly at a uniformly random time $\tau \in [T]$. We compute the average regret (in black), and 95% confidence interval for the mean (in shading) over 100 instances. Figure 7a and Figure 7c show the regret of the passive and active monitoring strategies, respectively, when the post-change preferences are unknown. Similarly, Figure 7b and Figure 7d present the regret of the passive and active monitoring strategies, respectively, when the post-change preferences are known.

We compare four algorithms tailored to abrupt changes in preferences. Algorithms 2 and 3 correspond to the risk-profile restrictions imposed in Section 5 and treat post-change preferences as unknown, so the retailer must both detect the change and relearn preferences after it occurs. By contrast, Algorithms 4 and 5, described in Appendix A.3, adopt the restrictive assumption that the retailer knows the post-change preferences and is uncertain only about the time of the change.

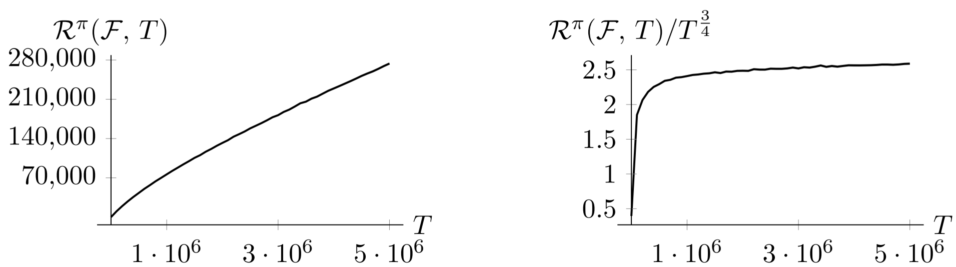
These policies still incur an unavoidable loss relative to a clairvoyant retailer because the change time is unknown, but they isolate the detection cost by removing the need to relearn preferences. Although this assumption is unrealistic, it defines a natural intermediate benchmark and clarifies how much revenue is lost from not knowing the post-change preferences, as well as how close our monitoring policies come to that idealized reference.

The change is passively detectable in this scenario because it affects the attraction parameters of products in the pre-change optimal assortment. As a result, Algorithms 3 and 5 are both applicable in this setting. To facilitate a fair comparison among these algorithms, we deliberately exclude the regret related to the learning phase of the preferences once the change is detected. This methodological choice allows us to isolate the opportunity cost of detecting preferences shifts and examine whether strategically leveraging the restrictions imposed on the class of admissible preferences' dynamics offers meaningful advantages for the retailer. Figure 7 illustrates the regret associated with each of these four algorithms.

Discussion. This scenario supports the value of tightening the risk profile. When the retailer recognizes that the change is passively detectable from its pre-change optimal assortment and restricts attention accordingly, it can monitor sales and intervene only when the data indicate a shift. In this regime, Algorithm 3, which does not assume knowledge of the post-change preferences, performs nearly as well as the idealized passive policy from Algorithm 5 that assumes full knowledge of these preferences; compare Figures 7a and 7b. Thus, conditional on passive detectability, the revenue loss from not knowing the post-change preferences is small. The effectiveness of Algorithm 3 depends on the separability parameter, which encodes the retailer's tolerance for small changes.

By contrast, if the retailer does not recognize that the change lies within the products it already offers and instead adopts the active monitoring strategy in Algorithm 2, then the regret increases sharply (Figure 7c). This increase is driven by the need to alternate between sentinel and steady cycles. In particular, after a change occurs, the policy may need to complete an ongoing steady cycle before starting a new sentinel cycle, hence delaying detection and inflating regret.

Finally, Algorithm 4 improves upon Algorithm 2 (Figures 7d versus 7c), highlighting the value of knowing the post-change preferences within an active-monitoring regime. This gain has two sources. First, Algorithm 2 must cycle through the set of test assortments, whereas Algorithm 4 can focus on a single informative assortment, reducing both monitoring cost and regret. Second, the statistical tests differ between the two procedures. Algorithm 2 relies on estimated purchase-probability gaps and therefore requires a relatively large sample size, whereas Algorithm 4 can rely on a likelihood ratio test. This heavier sampling requirement is reflected in Proposition 4, where the regret guarantee is stated only for sufficiently large time horizon T.



(a) Regret of Algorithm 1. (b) Regret of Algorithm 1 scaled by $T^{3/4}$.

Figure 8: Regret (and its scaled version) for the restart-based strategy $\pi(\mathcal{A}, \Delta)$ described in Algorithm 1. We compute the average regret (in black), and 95% confidence interval for the mean (imperceptible) over 100 instances.

Having shown that tightening the risk profile can unlock substantial performance gains, we now examine the opposite extreme, in which the retailer is overconservative and does not restrict the class of admissible preferences' dynamics to abrupt changes. In this case, the retailer adopts the restart-based strategy from Algorithm 1, with the dynamic assortment planning policy $\mathcal{A}$ from Agrawal et al. (2019), with Δ chosen as in Corollary 1. As depicted in Figure 8a, the regret incurred by this policy is higher than that of the strategies designed for abrupt changes, shown in Figure 7. Caution is warranted when comparing regret levels directly, since the opportunity cost of learning post-change preferences is omitted from the analyses of the monitoring-based strategies. Nonetheless, the regret of Algorithm 1 grows at rate $\mathcal{O}\left(T^{3/4}\right)$, as observed in Figure 8b. This rate is substantially larger than those in Figure 7, indicating that overconservatism can be costly over time.

Together with the seasonal diffusion scenario in Section 6.2, these results illustrate that poor performance under evolving preferences can come from two distinct sources. The first is inaction, when the retailer fails to plan for change at all. The second is overconservatism, when the retailer plans for a broader class of dynamics than necessary. Tightening the risk profile does not only reduce theoretical regret bounds; it also changes how the retailer should plan and respond to change. In particular, when the assortment itself provides sufficient coverage, robust planning does not require frequent restarts or active monitoring, but can rely on passive monitoring with targeted restarts. This heterogeneity highlights that robust planning is inherently retailer-specific: the same preferences' dynamics can induce very different opportunity costs depending on the structure and coverage of the assortments a retailer is able to offer.

7 Concluding remarks

Evolving customers' preferences can quietly erode a retailer's revenue over time. Much like in the boiling frog apologue, this operational risk may remain hidden unless it is actively recognized and anticipated. We show that retailers should exploit available frontline inputs to address this risk through upfront planning. By articulating a class of admissible preferences' dynamics against which robustness is sought, the retailer defines a risk profile that governs attainable performance and the structure of the appropriate strategy. Robustness to broad dynamics carries an unavoidable opportunity cost relative to a clairvoyant retailer. Yet, restart-based strategies, with the restart timing tied to the retailer's risk profile, provide a predictable and near-optimal way to control this cost.

At the same time, robustness need not imply rigidity. When the retailer is willing to trade some robustness for exploitable structure in preferences' dynamics, robust planning can move beyond periodic restarts toward monitoring sales data and detecting changes. In this regime, assortments serve not only as revenue-generating decisions but also as sensing instruments that reveal changes in preferences. Detectability then becomes central. When changes are passively detectable from the assortment a retailer already offers, adaptation can be achieved with logarithmic regret. When they are not, the cost rises sharply. As a result, two retailers with similar risk profiles may still experience different performance simply because their assortments differ in coverage and informational content.

The broader lesson is that managing evolving preferences is primarily a strategic planning problem rather than an algorithmic one. Balancing the remembering-forgetting trade-off can be reduced to articulating a risk profile that reflects the retailer's products, markets, and operational tolerance for change, and letting that choice govern when past information should be used and when it should be discarded. In this sense, robust planning is less about ever more complex dynamic assortment policies and more about deciding which preferences' dynamics are worth hedging against in the first place. At the same time, algorithmic development remains valuable. Designing assortments and monitoring schemes that jointly maximize revenue, improve the detectability of shifts in preferences, and preserve operational stability remains a promising direction for future research.

References

Agrawal, S., Avadhanula, V., Goyal, V., and Zeevi, A. (2019). "MNL-bandit: A dynamic learning approach to assortment selection". In: *Operations Research* 67.5, pp. 1453–1485.

Bernstein, F., Modaresi, S., and Sauré, D. (2019). "A dynamic clustering approach to data-driven assortment personalization". In: *Management Science* 65.5, pp. 2095–2115.

Besbes, O., Gur, Y., and Zeevi, A. (2015). "Non-stationary stochastic optimization". In: *Operations Research* 63.5, pp. 1227–1244.

Besbes, O. and Sauré, D. (Sept. 2014). "Dynamic Pricing Strategies in the Presence of Demand Shifts". In: *Manufacturing & Service Operations Management* 16, pp. 513–528.

Besbes, O. and Zeevi, A. (2009). "Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms". In: *Operations Research* 57.6, pp. 1407–1420.

— (2011). "On the minimax complexity of pricing in a changing environment". In: *Operations Research* 59.1, pp. 66–79.

Blanchet, J., Gallego, G., and Goyal, V. (2016). "A Markov chain approximation to choice modeling". In: *Operations Research* 64.4, pp. 886–905.

Caro, F. and Gallien, J. (2007). "Dynamic assortment with demand learning for seasonal consumer goods". In: *Management Science* 53.2, pp. 276–292.

Caro, F., Kök, A. G., and Martínez-de-Albéniz, V. (2020). "The future of retail operations". In: *Manufacturing & Service Operations Management* 22.1, pp. 47–58.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games.* Cambridge University press.

Döpper, H., MacKay, A., Miller, N., and Stiebale, J. (2024). *Rising markups and the role of consumer preferences.* Tech. rep. National Bureau of Economic Research.

Farias, V. F., Jagabathula, S., and Shah, D. (2013). "A nonparametric approach to modeling choice with limited data". In: *Management Science* 59.2, pp. 305–322.

Feldman, J. and Topaloglu, H. (2015). "Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits". In: *Production and Operations Management* 24.10, pp. 1598–1620.

Foster, D. P. and Vohra, R. (1999). "Regret in the online decision problem". In: *Games and Economic Behavior* 29.1-2, pp. 7–35.

Foussoul, A., Goyal, V., and Gupta, V. (2023). "Mnl-bandit in non-stationary environments". In: *arXiv preprint arXiv:2303.02504*.

Gallego, G. and Topaloglu, H. (2014). "Constrained assortment optimization for the nested logit model". In: *Management Science* 60.10, pp. 2583–2601.

Hampson, D. P. and McGoldrick, P. J. (2013). "A typology of adaptive shopping patterns in recession". In: *Journal of Business Research* 66.7, pp. 831–838.

Hannan, J. (1957). "Approximation to Bayes risk in repeated play". In: *Contributions to the Theory of Games* 3.2, pp. 97–139.

Hartmann, W. R. and Nair, H. S. (2010). "Retail competition and the dynamics of demand for tied goods". In: *Marketing Science* 29.2, pp. 366–386.

Honhon, D., Jonnalagedda, S., and Pan, X. A. (2012). "Optimal algorithms for assortment selection under ranking-based consumer choice models". In: *Manufacturing & Service Operations Management* 14.2, pp. 279–289.

Keskin, N. B. and Zeevi, A. (2017). "Chasing demand: Learning and earning in a changing environment". In: *Mathematics of Operations Research* 42.2, pp. 277–307.

Kök, A. G., Fisher, M. L., and Vaidyanathan, R. (2015). "Assortment planning: Review of literature and industry practice". In: *Retail Supply Chain Management: Quantitative Models and Empirical Studies*, pp. 175–236.

Lai, T. L. and Robbins, H. (1985). "Asymptotically efficient adaptive allocation rules". In: *Advances in Applied Mathematics* 6.1, pp. 4–22.

Li, S., Luo, Q., Huang, Z., and Shi, C. (2025). "Online Learning for Constrained Assortment Optimization Under Markov Chain Choice Model". In: *Operations Research* 73.1, pp. 109–138.

Lorden, G. (1971). "Procedures for reacting to a change in distribution". In: *The Annals of Mathematical Statistics*, pp. 1897–1908.

Mahajan, S. and Van Ryzin, G. (2001). "Stocking retail assortments under dynamic consumer substitution". In: *Operations Research* 49.3, pp. 334–351.

Rooderkerk, R. P., DeHoratius, N., and Musalem, A. (2022). "The past, present, and future of retail analytics: Insights from a survey of academic research and interviews with practitioners". In: *Production and Operations Management* 31.10, pp. 3727–3748.

Rusmevichientong, P., Shen, Z.-J. M., and Shmoys, D. B. (2010). "Dynamic assortment optimization with a multinomial logit choice model and capacity constraint". In: *Operations Research* 58.6, pp. 1666–1680.

Sauré, D. and Zeevi, A. (2013). "Optimal dynamic assortment planning with demand learning". In: *Manufacturing & Service Operations Management* 15.3, pp. 387–404.

Sheth, J. (2020). "Impact of Covid-19 on consumer behavior: Will the old habits return or die?" In: *Journal of Business Research* 117, pp. 280–283.

Sheth, J. N. (1967). "A review of buyer behavior". In: *Management Science* 13.12, B–718.

Talluri, K. and Van Ryzin, G. (2004). "Revenue management under a general discrete choice model of consumer behavior". In: *Management Science* 50.1, pp. 15–33.

Thomas, M. and Joy, A. T. (2006). *Elements of information theory*. Wiley-Interscience.

Ton, Z. (2014). *The good jobs strategy: How the smartest companies invest in employees to lower costs and boost profits*. Houghton Mifflin Harcourt.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University press.

Tsybakov, A. B. (2003). *Introduction à l'estimation non paramétrique*. Vol. 41. Springer Science & Business Media.

Van Ryzin, G. and Vulcano, G. (2015). "A market discovery algorithm to estimate a general class of nonparametric choice models". In: *Management Science* 61.2, pp. 281–300.

Wei, C.-Y. and Luo, H. (2021). "Non-stationary reinforcement learning without prior knowledge: An optimal black-box approach". In: *Conference on Learning Theory*. PMLR, pp. 4300–4354.

Online Appendix

(Saving Kermit: Dynamic Assortment Planning in a Boiling Market)

A.1 Constructing a set of test assortments to detect changes

Constructing the set of test assortments $\mathcal{E}$ is critical to balance the remembering-forgetting trade-off in Algorithm 2. We propose an approach to construct $\mathcal{E}$ that exploits the separability condition. Our construction relies on two key assumptions about the underlying preferences:

(i) The pre- and post-change distributions F^1 and F^τ are both parametric, each with corresponding purchasing probability given by $p(\cdot, \theta)$ for some parameters θ.

(ii) For any $\rho \in \mathbb{R}_+^N$ with $\sum_{i \in \mathcal{N}} \rho_i < 1$, there exists a unique parameter vector $\eta(\rho)$ such that $p_i(S, \eta(\rho)) = \rho_i$ for all $i \in S$ and $S \in \mathcal{S}$.

Assumption (i) applies to parametric models commonly used in assortment planning, such as MNL, though it excludes ranking-based ones. Assumption (ii), an *identifiability* condition, ensures that model's parameters can be uniquely inferred from estimated purchasing probabilities. The identifiability condition is naturally satisfied by the MNL choice model (Sauré and Zeevi 2013).

To construct $\mathcal{E}$, we partition the product set $\mathcal{N}$ into $\lceil N/K \rceil$ disjoint subsets A_j, each of size at most K, and define $\mathcal{E} := \{A_1, \ldots, A_{\lceil N/K \rceil}\}$. We refer to this construction as the *partitioning approach*. In the worst-case, the regret bound from Proposition 2 scales with $|\mathcal{E}|$, growing as $\mathcal{O}\big(\binom{N}{K}\big)$ in the complete absence of structural assumptions on the customers' preferences. Assumptions (i) and (ii) ensure that any change in preferences is reflected in the model's parameters. By following the partitioning approach, we can exploit the sales data to construct an estimator for the model's parameters, compare this estimator to the pre-change parameters, and detect changes in preferences when they occur. Therefore, partitioning reduces this complexity to $\mathcal{O}\big(\lceil N/K \rceil\big)$.

Note that the partitioning approach does not guarantee that $\mathcal{E}$ includes an assortment that satisfies the separability condition within the class $\tilde{\mathcal{F}}_U$. However, one could modify the statistical test for change detection in Algorithm 2 by:

$$\max \left\{ \left| p_i(S, \hat{F}) - p_i(S, F^1) \right| \, : \, i \in S \cup \{0\} \, , \, S \in \mathcal{S} \right\} > \kappa/2,$$

where $\hat{F}$ denotes the preferences estimated from the purchasing data collected by offering the assortments from $\mathcal{E}$. This modification comes at the expense of an increase in the computational complexity. Therefore, for the purposes of this study, we assume that the partitioning approach returns an assortment in $\mathcal{E}$ that satisfies the separability condition within class $\tilde{\mathcal{F}}_U$.

A.2 Implementation details of Section 6

This section outlines the setup used for the case studies from Sections 6.2 and 6.3, including our parameter choices and some adaptations of our algorithms to handle limitations from the data.

Section 6.2. We implement Algorithm 1 using the policy by Agrawal et al. (2019) as a subroutine for learning static preferences. To calibrate our restart-based strategy, we fix the value taken by the magnitude of the class $\mathcal{F}$ to $\mathcal{M}(\mathcal{F}, T) = 5 \times 10^{-2}$. Following Corollary 1, we use a sub-segment size of $\Delta = \lceil T^{1/2} \cdot \mathcal{M}(\mathcal{F}, T)^{1/2} \rceil$, as an input to our restart-based strategy.

Section 6.3. To isolate the cost of monitoring, we exclude the regret component due to estimating post-change preferences. Rather than relying on empirical averages, we compute regret as the gap in *expected* revenue between the clairvoyant retailer (i.e., the oracle) and the implemented policy. For false alarms, when a change is detected prematurely, we compute the one period regret as $r(S^*(\nu^1), \nu^1) - r(S^*(\nu^\tau), \nu^1)$. This calculation serves as a proxy for the cost of learning the wrong regime, in the sense that the policy behaves as if it should learn the post-change preferences while the actual preferences remain governed by ν^1. We address the omitted learning term explicitly in our theoretical results and in the main body of the paper.

Policies based on the change-detection approach introduced in this paper partition the horizon into sub-segments devoted to either sentinel cycles or steady cycles, or to a combination of both. The optimal size of these segments depends on the similarity between the pre- and post-change preferences. Loosely speaking, smaller changes require longer sentinel cycles to ensure reliable detection (we refer to the corresponding sections and proofs for further details). In our setting, the pre- and post-change preferences, obtained by calibrating MNL models on the dataset, are close to each other. This closeness results in a low KL divergence between the preferences of order $\mathcal{O}(10^{-3})$, which, in turn, necessitates the use of large sub-segments for sentinel cycles of order at least $\mathcal{O}(10^6)$.

For the active monitoring strategy (Algorithm 2), we set κ to the infinite-norm between the pre- and post-change preferences conditional on the assortment $S^*(\nu^1)$. This modification improves the stability of the statistical test used within the procedure. We use sub-segment sizes $D_1\sqrt{T}$ for steady cycles and $D_2 \log T$ for sentinel cycles, with $D_1 = 100$ and $D_2 = 5{,}000$. We construct the set of test assortments using the partitioning approach from the Appendix A.1. The same principle applies for the passive monitoring strategy (Algorithm 3). We set $\kappa = \varepsilon$ and use sub-segment sizes $D_2 \log T$ with $D_2 = 5{,}000$. For both Algorithm 4 and Algorithm 5, we adopt the same sub-segment sizes as for the setting with unknown post-change preferences.

A.3 Intermediate benchmark for monitoring

In this section, we discuss a setting similar to the one from Section 5 in which the retailer seeks robustness against abruptly changing preferences. However, in what follows, the retailer specifies the post-change preferences so that only the time of the change remains unknown. Similar assumptions have been studied in the context of pricing under dynamic demand models (Besbes and Zeevi 2011; Besbes and Sauré 2014). Our goal, first, is to understand how access to post-change preferences can aid the retailer in designing more effective robust planning strategies. Furthermore, this setting serves as an intermediate benchmark for comparison with the more challenging case presented in Section 5, where post-change preferences are assumed to be unknown to the retailer.

Since the retailer is assumed to know both the pre- and post-change preferences (when a change occurs), the preferences $F^{(\mathbb{N})} \in \mathcal{F}_A$ are fully characterized by the pair (F^1, F^τ) and the change time τ. Throughout this section, we refer to F^1 and F^τ, always assuming (implicitly) that there exists some preferences $F^{(\mathbb{N})} \in \mathcal{F}_A$ that satisfy $F^t = F^1$ for all $t < \tau$ and $F^t = F^\tau$ for all $t \geq \tau$. We also denote by $\mathcal{F}_A(F^1, F^\tau)$ the subset of sequences in $\mathcal{F}_A$ with pre- and post-change preferences given by F^1 and F^τ, respectively. We adopt the same assumptions and notational conventions introduced in Section 5.1. Additionally, we assume that $\mathcal{F}_A$ is such that the following quantity:

$$\vartheta \equiv \vartheta(\mathcal{F}_A) := \sup\left\{ \left| \log p_i(S, F^1) - \log p_i(S, F^\tau) \right| \,:\, \forall\, i \in S \cup \{0\},\, S \in \mathcal{S},\, F^{(\mathbb{N})} \in \mathcal{F}_A \right\},$$

satisfies $\vartheta < \infty$ so that the magnitude of the class $\mathcal{F}_A$ is bounded above uniformly. This assumption is equivalent to assuming that the probability of purchase for any product cannot be made arbitrarily small by an adversary environment.

A.3.1 Active monitoring with known post-change preferences

We consider passively undetectable changes as first discussed in Section 5.2. That is, we assume that preferences $F^{(\mathbb{N})}$ belong to $\mathcal{F}_U$, so that the pre- and post-change preferences F^1 and F^τ cannot be distinguished by only offering the pre-change optimal assortment $S^*(F^1)$.

A.3.1.1 A fundamental limit on the achievable performance

We establish a lower bound on the regret that any assortment strategy must incur. The proof follows the same line of reasoning as in the setting with unknown post-change preferences and is based on a constructive change-point argument. Specifically, we distinguish between assortment strategies that are guaranteed to sufficiently forget and those that fail to do so. For each case, a change point is constructed in an adversarial manner to establish the lower bound on the regret.

Proposition 5. *There exists some finite constant $C \equiv C(\gamma, \vartheta) > 0$, such that, for $T \geq 2$:*

$$\widetilde{\mathcal{R}}\big(\mathcal{F}_U, T\big) \geq C\sqrt{T}.$$

The lower bound in Proposition 5 shows that knowledge of the post-change preferences does not alter the regret's dependence on the number of customers T. The regret remains of the same order as in Proposition 1, where the post-change preferences are unknown, and the difference between the two results lies in the constant in front of the term $\sqrt{T}$.

A.3.1.2 An improved active monitoring strategy

In Algorithm 4, we specialize the active monitoring strategy from Algorithm 2 to the setting in which the post-change preferences F^τ are known. This specialization manifests in two key aspects. First, the algorithm leverages F^τ by performing a log-likelihood ratio test on the data collected from a single test assortment to determine whether the observed data is more likely to have been generated by F^1 or F^τ, conditional on a given assortment $S \in \mathcal{S}$. Second, once a change is detected, the algorithm immediately switches to offering the post-change optimal assortment. The only requirement we impose on the test assortment S is that it discriminates F^τ from F^1.

Algorithm 4 Active monitoring strategy $\pi\left(D, F^1, F^\tau, S\right)$

Input: A constant $D > 0$, two distributions F^1 and F^τ, and a test assortment S
Initialize: Set $detect = False$, $t = 0$, $\Delta_o := D\sqrt{T}$, $\Delta_e := D\log T$
while $detect = False$ and $t \leq T$ **do**
  Offer $S^u = S^*(F^1)$ for $u = t+1, \ldots, t+\Delta_o$ (sentinel cycle)
  Offer assortment S to Δ_e customers (steady cycle)
  if $\sum_{u=t+\Delta_o+1}^{t+\Delta_o+\Delta_e} \log p_{i^u}(S^u, F^1) - \log p_{i^u}(S^u, F^\tau) < 0$ **then**
    $detect = True$ (change detected)
  $t = t + \Delta_o + \Delta_e$
Offer $S^*(F^\tau)$ to customers $t+1, \ldots, T$ (post-change policy)

The constant D, used as an input to Algorithm 4, can be determined by specifying a vector $\alpha = (\alpha_{\mathrm{I}}, \alpha_{\mathrm{II}})$, in addition to the assortment S used within the policy. The vector α encodes the desired Type I and Type II error levels for the statistical test employed in the change detection step. The computation of $D \equiv D(\alpha, S)$ is detailed in the proof of Proposition 6, which establishes an upper bound on the regret of Algorithm 4, provided that the preferences F^1 and F^τ are distinguishable under the assortment S. Specifically, the assortment S, which we refer to as a *test assortment*, must be chosen such that the pre- and post-change preferences differ when conditioned on S.

Proposition 6. *For $\alpha := (\alpha_I, \alpha_{II})$, $S \in \mathcal{S}$, $D \equiv D(\alpha, S)$, let $\pi \equiv \pi(D, F^1, F^\tau, S)$ be the policy defined in Algorithm 4. Then, there exists a constant $C \equiv C(\alpha, S) > 0$, such that, for $T \geq 2$:*

$$\mathcal{R}^\pi\big(\mathcal{F}_U(F^1, F^\tau)\big) \leq C \log T \sqrt{T}.$$

The upper bound on regret in Proposition 6 aligns with the lower bound in Proposition 5, differing only by a logarithmic factor. In other words, Algorithm 4 achieves near-optimal performance. However, the choice of the assortment S plays a central role in the policy's effectiveness to balance the remembering-forgetting trade-off. An inappropriate choice may lead to a high value of C in Proposition 6 and therefore to higher regret in the worst case.

The principal advantage of knowing the post-change preferences F^τ lies in its effect on the constant terms in the regret bound. In this case, the policy can forgo the costly hypothesis testing procedure that would otherwise require offering a large number of assortments to collect sufficient purchasing data. Instead, it suffices to offer a single well chosen assortment.

A.3.2 Passive monitoring with known post-change preferences

Next, we consider a setting in which the retailer seeks robustness against abrupt changes when the pre-change optimal assortment provides sufficient coverage for the change to be passively detectable, that is, when $F^{(\mathbb{N})} \in \mathcal{F}_D$. Accordingly, F^1 and F^τ can be distinguished based on the sales data collected by offering the pre-change optimal assortment $S^*(F^1)$. However, in contrast to Section 5.3, we assume that the retailer knows the post-change preferences F^τ.

A.3.2.1 A fundamental limit on the achievable performance

We now derive a fundamental performance bound for any assortment strategy when post-change preferences are assumed to be known and passively detectable.

Proposition 7. *There exists constants $C \equiv C\left(\gamma, \vartheta\right) > 0$ and $t \equiv t\left(\gamma, \vartheta\right) > 0$, such that, for $T \geq t$:*

$$\widetilde{\mathcal{R}}^*\left(\mathcal{F}_D, T\right) \geq C \log T.$$

Proposition 7 establishes a lower bound on the regret for the setting in which post-change preferences are known. This result parallels Besbes and Zeevi (2011), who study dynamic pricing under abrupt demand shifts with known post-change demand. In particular, the game-theoretic interpretation of Proposition 3 remains valid here, since the underlying proofs are essentially identical.

A.3.3 Passive monitoring with known post-change preferences

Next, we present a variation of the passive monitoring strategy from Algorithm 3 which makes use of the post-change preferences. Yet, the core idea remains the same: the retailer initially offers the pre-change optimal assortment to passively monitor sales data. However, as F^τ is known in this setting, we use a log-likelihood ratio test for change detection. If a change in preferences is detected, then the retailer offers the post-change optimal assortment. In contrast to Algorithm 3, there is no need to relearn preferences as they are assumed to be known in the present setting.

Algorithm 5 Passive monitoring strategy $\pi\left(D, F^1, F^\tau\right)$

Input: A constant $D > 0$, two distributions F^1 and F^τ, respectively
Initialize: Set $detect = False$, $t = 0$, $\Delta = D\log(T)$
while $detect = False$ and $t < T$ **do**
  Offer $S^u = S^*\left(F^1\right)$ for $u = t+1, \ldots, t+\Delta$ (passive monitoring)
  if $\sum_{u=t+1}^{t+\Delta} \log p_{i_u}(S^*\left(F^1\right), F^1) - \log p_{i_u}(S^*\left(F^1\right), F^\tau) < 0$ **then**
    $detect = True$ (change detected)
  Set $t = t + \Delta$
Offer $S^t = S^*(F^\tau)$ to customers $t+1, \ldots, T$ (post-change policy)

The constant D, required as an input for Algorithm 5, can be naturally determined through the selection of two parameters α_{I} and α_{II}, which specifies the Type I and II errors level for the log-likelihood test used in the policy. The detailed derivation of this constant can be found in the proof of Proposition 8 which provides an upper bound for the regret of Algorithm 5.

Proposition 8. *For $\alpha = (\alpha_I, \alpha_{II}) \in (0,1)^2$ and $D \equiv D(\alpha) > 0$, let $\pi \equiv \pi\left(D, F^1, F^\tau\right)$ be the policy defined in Algorithm 5. Then, there exist finite constants $C_1 \equiv C_1(\alpha_I) > 0$ and $C_2 \equiv C_2(\alpha_{II}) > 0$, such that, for $T \geq 2$:*

$$\mathcal{R}^\pi\left(\mathcal{F}_D(F^1, F^\tau)\right) \leq C_1 + C_2 \log T.$$

The upper bound above matches the lower bound from Proposition 7 up to a constant term. Moreover, if a policy $\mathcal{A}$ designed to learn static preferences achieves a regret of order $\mathcal{O}(\log T)$, then the upper bound from Proposition 4 for the case in which the post-change preferences are unknown remains of the same order as the bound in Proposition 8. In that regime, the main driver of regret is the unknown change time rather than uncertainty about the post-change preferences. This observation aligns with the previous discussion in Appendix A.3.1.

References

Agrawal, S., Avadhanula, V., Goyal, V., and Zeevi, A. (2019). "MNL-bandit: A dynamic learning approach to assortment selection". In: *Operations Research* 67.5, pp. 1453–1485.

Besbes, O. and Sauré, D. (Sept. 2014). "Dynamic Pricing Strategies in the Presence of Demand Shifts". In: *Manufacturing & Service Operations Management* 16, pp. 513–528.

Besbes, O. and Zeevi, A. (2011). "On the minimax complexity of pricing in a changing environment". In: *Operations Research* 59.1, pp. 66–79.

Sauré, D. and Zeevi, A. (2013). "Optimal dynamic assortment planning with demand learning". In: *Manufacturing & Service Operations Management* 15.3, pp. 387–404.

Electronic Companion - technical proofs

(Saving Kermit: Dynamic Assortment Planning in a Boiling Market)

Notations. Let F and G be probability distributions defined on a common discrete probability space $(\Omega, \mathcal{B}, \mathbb{P})$. The KL divergence between F and G is defined as follows (Thomas and Joy 2006):

$$\mathcal{K}(F, G) := \sum_{\omega \in \Omega} F(\omega) \log \frac{F(\omega)}{G(\omega)}.$$

In particular, throughout this work, we often refer to the KL divergence between two distributions conditional on an event S. Correspondingly, given some event S we denote by $\mathcal{K}(F, G\,;\, S)$ the the KL divergence between the conditional distributions $F(\cdot \mid S)$ and $G(\cdot \mid S)$. Moreover, to measure the variability of a given sequence of customers' preferences $F^{(\mathbb{N})} \in \mathcal{F}$, we use the notation:

$$\mathcal{K}^t(S) = \sum_{i \in S \cup \{0\}} p_i(S, F^t)\big(\log p_i(S, F^t) - \log p_i(S, F^{t-1})\big),$$

as originally introduced in Section 4.

The infinity norm is denoted by $\|\cdot\|_\infty$. Expectations taken with respect to the probability measure $\mathbb{P}$ are denoted explicitly as $\mathbb{E}_\mathbb{P}$. Some statements may be understood as holding *almost surely* (i.e., with probability 1 under the appropriate probability measure), although we omit explicit references for notational simplicity. We write $a_n = o(b_n)$ to mean that $a_n/b_n \to 0$ as $n \to +\infty$, and $a_n = \mathcal{O}(b_n)$ if there exists a constant $C > 0$ such that $|a_n| \leq C|b_n|$ for sufficiently large n. The indicator function $\mathbf{1}(\cdot)$ takes the value 1 if and only if its argument is true. Throughout the proofs, the terms *environment* and *nature* are used interchangeably to refer to an adversary that selects the preferences $F^{(\mathbb{N})}$ as originally discussed in Section 3 .

E.C.1 Proofs for Section 4

We present detailed proofs of the theoretical results established in Section 4. Our analysis is conducted under the fundamental assumption that customers' preferences evolve over time, with changes bounded according to the magnitude $\mathcal{M}(\mathcal{F}, T)$, as formally introduced and discussed in Section 4.1. We begin by rigorously establishing a lower bound on the regret that any admissible policy can achieve, as stated in Theorem 1. We then derive an upper bound on the regret incurred by our proposed restart-based strategy, thereby proving the performance guarantee stated in Theorem 2. Both results appear in Sections 4.2 and 4.3, respectively.

Proof of Theorem 1. For $T \geq 2$ be the time horizon. We define $M_T \equiv \mathcal{M}(\mathcal{F}, T)$ as the magnitude of the class of admissible preferences' dynamics $\mathcal{F}$. Moreover, if $T := o(M_T)$, then one can construct an instance for which no policy can achieve a sub-linear regret. Therefore, without loss of generality, we assume that $1 < M_T < \frac{1}{4}T$. In addition, we assume that the profit vector $\mathbf{w} \equiv (w_i : i \in \mathcal{N})$ satisfies $w_i = 1$ for all $i \in \mathcal{N}$. This assumption entails no loss of generality, since the profit terms in the regret can be lower bounded by $\min\{w_i : i \in [N]\}$.

We partition the selling horizon $[T]$ into sub-segments of size $\Delta \in [T]$, defined as $\Delta := \lfloor T^{1/2} M_T^{-1/2} \rfloor$. Let $\tilde{T}-1 := \lceil T/\Delta \rceil - 1$ denote the number of sub-segments, denoted by $\mathcal{F}_1, \ldots, \mathcal{F}_{\tilde{T}-1}$. Each sub-segment has cardinality Δ, except possibly $\mathcal{F}_{\tilde{T}-1}$, which may be smaller. For simplicity, and without loss of generality, we fix the number of products in each assortment to $K = 1$. Our construction can be extended to general K but becomes more technically involved. We then fix an arbitrary admissible policy $\pi \equiv (\psi_t(\mathcal{H}_{t-1}))_{t=1}^{T} \in \mathcal{P}$, where ψ_t maps the past history to an assortment from $\mathcal{S}$. To simplify notation, we omit the explicit dependence of π on the filtration $(\mathcal{H}_t)_{t=0}^{T-1}$.

We establish a lower bound on the achievable regret for any arbitrary policy π through a constructive approach. Specifically, we demonstrate this bound by constructing an adversarial instance that forces any policy to incur the minimal regret stated in the proposition. To formalize this result, we first define a subset of preferences as follows:

$$\mathcal{M}' := \left\{ F^{(\mathbb{N})} \;:\; F^t \in \{F_a, F_b\},\, F^t = F^{t-1} \;\forall\, t \notin \{\Delta, 2\Delta, \ldots, \tilde{T}\Delta\} \right\},$$

where F_a and F_b are two MNL choice models with attraction parameters given by the $(N+1)$-dimensional vectors ν^a, and ν^b, respectively, which, in turn, are defined as follows:

$$\nu^a := \left(1, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{4}, \frac{1}{2} - \zeta\right) \quad \text{and} \quad \nu^b := \left(1, \frac{1}{2}, \frac{1}{4}, \ldots, \frac{1}{4}, \frac{1}{2} + \zeta\right),$$

where $\zeta := \frac{1}{4}(M_T/\tilde{T})^{\frac{1}{2}} < \frac{1}{4}$. Also, $\nu_0^a = \nu_0^b = 1$ are the parameters for the non-purchase decision.

Accordingly, we obtain the following lower bound on the difference in expected revenue for a single sale between the assortment $\{N\}$ and any other assortment $S \in \mathcal{S}$, with S different from $\{N\}$:

$$r(\{N\}, F_b) - r(S, F_b) \geq \min\left\{p_N(\{N\}, F_b) - p_i(S, F_b) \;:\; i \in [N-1]\right\} \geq \zeta. \tag{E.C.1}$$

A similar bound holds for $r(\{1\}, F_a) - r(S, F_a)$ when S is different from $\{1\}$.

Step 1 ($\mathcal{M}'$ is well-defined). To begin, we show that customers' preferences belonging to $\mathcal{M}'$ have a cumulative variability bounded above by M_T. By assumption, each feasible assortment has cardinality $K = 1$, and hence, $\mathcal{S} \equiv \big\{\{i\} : i \in [N]\big\}$. Accordingly, by the definition of the attraction parameters ν^a and ν^b, one can verify that for all feasible assortments $S \neq \{N\}$, the KL divergence satisfies $\mathcal{K}\left(F_a, F_b\,;\, S\right) = 0$. Indeed, conditional on $S \neq \{N\}$, the distributions F_a and F_b coincide.

However, conditional on the assortment $S = \{N\}$, we have that $\mathcal{K}\left(F_a, F_b\,;\,S\right) \neq 0$ and that the following sequence of inequalities holds:

$$\mathcal{K}(F_a, F_b\,;\,\{N\}) = \mathcal{K}(F_b, F_a\,;\,\{N\}) = 2\zeta \log\big(\frac{1+2\zeta}{1-2\zeta}\big) \leq 2\zeta\big(\frac{1+2\zeta}{1-2\zeta} - 1\big) = \frac{8\zeta^2}{1-2\zeta} \overset{(a)}{\leq} 16\zeta^2 \overset{(b)}{\leq} M_T/\tilde{T},$$

where (a) follows from that $\zeta \leq \frac{1}{4}$, and (b) follows from the definition of ζ.

Next, we fix some preferences $F^{(\mathbb{N})} \in \mathcal{M}'$ arbitrarily, where the customers' preferences on each sub-segment $\mathcal{F}_1$ to $\mathcal{F}_{\tilde{T}-1}$ are either F_a or F_b with probability 1/2 each. Therefore, since $F^{(\mathbb{N})} \in \mathcal{M}'$, we obtain the following upper bound on the variability of theses preferences:

$$\sum_{t=2}^{T} \max\left\{\mathcal{K}^t(S)\,:\,S \in \mathcal{S}\right\} \leq \sum_{j=1}^{\tilde{T}-1} \frac{M_T}{\tilde{T}} \leq M_T.$$

Thus, preferences $F^{(\mathbb{N})} \in \mathcal{M}'$ are guaranteed to have a variability that is bounded above by M_T.

Step 2 (Measuring the deviation between scenarios). We fix two preferences' dynamics $F, G \in \mathcal{M}'$ arbitrarily. Then, we denote by $\mathbb{P}_F^{\pi,\mathcal{F}_j}$ the probability distribution of customer's purchase decisions within the sub-segment $\mathcal{F}_j$, where $j \in [\tilde{T}-1]$, whenever the preferences are given by F and the assortment policy is π. Next, we introduce Z^j, the random vector that corresponds to the customer's purchase decisions within sub-segment $\mathcal{F}_j$.

We define $z^j \in \{0,1\}^{|\mathcal{F}_j| \times N}$, such that $z^j_{t,i} = 0$, if $i \notin \psi_t$, and we derive a closed-form formula for the probability distribution of customer's purchase decisions. The following equality holds:

$$\mathbb{P}_F^{\pi,\mathcal{F}_j}\left[Z^j = z^j\right] = \prod_{t \in \mathcal{F}_j} \mathbb{P}_F^{\pi,\mathcal{F}_j}\left[Z^j_t = z^j_t\right] = \prod_{t \in \mathcal{F}_j} F^t(z^j \mid \psi_t),$$

where similar observation remains valid whenever the preferences' dynamic F is replaced by G.

The closed-form formula for the probability distributions of customer's purchase decisions is used to measure how these two scenarios differ from each others. In particular, we compute the KL divergence between distributions that are induced by the two preferences' dynamics F and G. Formally:

$$\begin{aligned}
\mathcal{K}\big(\mathbb{P}_F^{\pi,\mathcal{F}_j}, \mathbb{P}_G^{\pi,\mathcal{F}_j}\big) &:= \mathbb{E}_{\mathbb{P}_F^{\pi,\mathcal{F}_j}}\big[\log\big(\frac{\mathbb{P}_F^{\pi,\mathcal{F}_j}[Z]}{\mathbb{P}_G^{\pi,\mathcal{F}_j}[Z]}\big)\big] \\
&= \mathbb{E}_{\mathbb{P}_F^{\pi,\mathcal{F}_j}}\big[\log\big(\frac{\prod_{t \in \mathcal{F}_j} F^t(Z^t \mid \psi_t)}{\prod_{t \in \mathcal{F}_j} G^t(Z^t \mid \psi_t)}\big)\big] \\
&= \mathbb{E}_{\mathbb{P}_F^{\pi,\mathcal{F}_j}}\big[\sum_{t \in \mathcal{F}_j} \log\big(\frac{F^t(Z^t \mid \psi_t)}{G^t(Z^t \mid \psi_t)}\big)\big] \\
&= \sum_{t \in \mathcal{F}_j} \mathbb{E}_{F^t}\big[\log\big(\frac{F^t(Z^t \mid \psi_t)}{G^t(Z^t \mid \psi_t)}\big)\big]
\end{aligned}$$

$$\overset{(a)}{=} \sum_{t\in\mathcal{F}_j} \mathbb{E}_{F^t}\big[\log\big(\frac{F^t(Z^t \mid \{N\})}{G^t(Z^t \mid \{N\})}\big)\mathbf{1}\big(\psi_t = \{N\}\big)\big]$$

$$\overset{(b)}{\leq} 2\zeta\log\big(\frac{1+2\zeta}{1-2\zeta}\big)\sum_{t\in\mathcal{F}_j}\mathbb{E}_{F^t}\big[\mathbf{1}\big(\psi_t = \{N\}\big)\big]$$

$$\leq 2\zeta\big(\frac{1+2\zeta}{1-2\zeta}-1\big)\Delta \leq \frac{8\zeta^2}{1-2\zeta}\Delta \overset{(c)}{\leq} 16\zeta^2\Delta \overset{(d)}{\leq} M_T\Delta^2/T \leq 1 \equiv \beta,$$

where (a) follows from that $F, G \in \mathcal{M}'$, and the definition of F_a and F_b. Then, (b) follows from the definition of ν^a and ν^b. Also, (c) holds since $\zeta := \frac{1}{4}(M_T/\tilde{T})^{\frac{1}{2}} < \frac{1}{4}$. Moreover, (d) follows from the definition of ζ. Finally, we define $\beta \in \mathbb{R}$ as $\beta = 1$ throughout the remainder of the proof.

Step 3 (Hypothesis test and Tsybakov's technique). We derive a lower bound on the exploration frequency of the policy π. First, we fix some sub-segment index $j \in [\tilde{T}-1]$ arbitrarily. From Step 2, we know that $\mathcal{K}(\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}, \mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}) \leq \beta$, where $\mathbf{F}_a := (F_a, \ldots, F_a)$, and $\mathbf{F}_b := (F_b, \ldots, F_b)$. Finally, given that $\mathcal{F}_j := \{\ell_j, \ldots, \ell_{j+1}-1\}$, we consider the following hypotheses test:

$$H_0: \quad Z^t \sim \mathbf{F}_a,\ t \in \mathcal{F}_j,$$

$$H_1: \quad Z^t \sim \mathbf{F}_b,\ t \in \mathcal{F}_j.$$

Let ϕ be any decision rule from the set of assortment and customer's purchase decisions in $\mathcal{F}_j$ into $\{0,1\}$. By convention, $\phi = 0$ indicates that the null hypothesis H_0 is not rejected, and $\phi = 1$ implies that the null hypothesis is rejected. Then, if H_0 is true, then the random vector Z is $\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}$-distributed, and if H_1 is true, then the random vector Z is $\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}$-distributed.

Next, we leverage Theorem 2.2 by Tsybakov (2003), to derive a lower bound for the probability of the Type I or II errors. Specifically, we obtain the following inequality:

$$\inf_{\phi}\max\big\{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}\big[\phi \neq 0\big], \mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}\big[\phi \neq 1\big]\big\} \geq \max\big\{\frac{1}{4}\exp(-\beta), \frac{1-\sqrt{\beta/2}}{2}\big\}.$$

Accordingly, by taking the complementary event, we obtain the following lower bound:

$$\inf_{\phi}\min\big\{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}\big[\phi = 0\big], \mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}\big[\phi = 1\big]\big\} \geq \max\big\{\frac{1}{4}\exp(-\beta), \frac{1-\sqrt{\beta/2}}{2}\big\}.$$

Next, we construct a decision rule ϕ that is based on the decision from policy π. Formally:

$$\phi(\pi) = \begin{cases} 0 & \text{if } \sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t \neq \{N\}\big) \leq \Delta/2, \\ 1 & \text{if } \sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t \neq \{N\}\big) > \Delta/2, \end{cases}$$

where ϕ depends on the observed realization of the purchase decision through the filtration $(\mathcal{H}_t)_{t=0}^{\ell_j-1}$.

We now apply the previously established lower bound for admissible decision rules to the one that is induced by the policy π. The analysis bifurcates into two cases, depending on whether the

Type I or Type II error exhibits higher probability. Specifically:

Case 1. To begin, we assume that $\min\big\{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}\big[\phi=0\big],\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}\big[\phi=1\big]\big\}=\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}\big[\phi=1\big]$. Then, we derive an upper bound on the number of time policy π does not select the optimal assortment (if the customer purchases are $\mathbf{F}_b$-distributed). Formally, we obtain the following upper bound:

$$\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t\neq\{N\}\big)\geq\frac{1}{2}\Delta\big]\leq 2\Delta^{-1}\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t\neq\{N\}\big)\big],$$

where we use the Markov's inequality to obtain the inequality (Jacod and Protter 2012).

We derive a lower bound on the expected frequency that π picks an assortment that is not optimal whenever the customer purchases are $\mathbf{F}_b$-distributed. Thus, the following inequality holds:

$$\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t\neq\{N\}\big)\big]\geq\frac{1}{8}\exp\left(-\beta\right)\Delta=\frac{1}{8}\exp\left(-1\right)\Delta.$$

Case 2. Next, we consider the case, where $\min\{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}\big[\phi=0\big],\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}\big[\phi=1\big]\}=\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}[\phi=0]$. Then, assume that $\phi=0$. As a consequence, the following inequality holds:

$$\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t=\{N\}\big)\geq\frac{1}{2}\Delta.$$

We hence obtain the following inequality:

$$\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t=\{N\}\big)\geq\frac{1}{2}\Delta\big]\leq 2\Delta^{-1}\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t=\{N\}\big)\big],$$

where we use the Markov's inequality to obtain the inequality.

We derive a lower bound on the expected frequency that π picks an assortment that is optimal with respect to F_b whenever the customer purchases are $\mathbf{F}_a$-distributed. That is:

$$\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t=\{N\}\big)\big]\geq\frac{1}{8}\exp\left(-1\right)\Delta.$$

Step 4 (Lower bound for the regret). Assume that, for each $j\in[\tilde{T}-1]$, nature selects either F_a or F_b as the customers' preferences in sub-segment $\mathcal{F}_j$, with probability $\frac{1}{2}$. The resulting preferences $F^{(\mathbb{N})}$ thus belong to $\mathcal{M}'$. Accordingly, we derive the following lower bound on the difference between the expected revenue of the oracle and that achieved by any policy π:

$$\begin{aligned}J^*(F^{(\mathbb{N})},T)-J^\pi(F^{(\mathbb{N})},T)\geq&\sum_{j\in[\tilde{T}-1]}\big[\frac{1}{2}\big(\Delta r(\{1\},F_a)-\sum_{t\in\mathcal{F}_j}r(\psi_t,F_a)\big)\\&+\frac{1}{2}\big(\Delta r(\{N\},F_b)-\sum_{t\in\mathcal{F}_j}r(\psi_t,F_b)\big)\big]\\\overset{(a)}{\geq}&\sum_{j\in[\tilde{T}-1]}\frac{\zeta}{2}\big(\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t\neq\{1\}\big)\big]+\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t\neq\{N\}\big)\big]\big)\end{aligned}$$

$$\geq \frac{1}{16}\exp\left(-1\right)\zeta(\tilde{T}-1)\Delta \geq \frac{\sqrt{2}-1}{(16)^2\sqrt{2}}\exp\left(-1\right)T^{\frac{3}{4}}M_T^{\frac{1}{4}},$$

where (a) follows from (E.C.1).

In particular, as this lower bound is valid for any admissible policy π, which imply

$$\mathcal{R}^*\left(\mathcal{F},T\right) \geq \frac{1}{16}\exp\left(-1\right)\zeta(\tilde{T}-1)\Delta \geq \frac{\sqrt{2}-1}{(16)^2\sqrt{2}}\exp\left(-1\right)T^{\frac{3}{4}}M_T^{\frac{1}{4}},$$

which holds by the definition of regret. ■

We now proceed to establish an upper bound on the regret associated with the restart-based strategy as formally described in Algorithm 1.

Proof of Theorem 2. For $\mathcal{A} \in \mathcal{P}$ and $\Delta \leq T$, let $\pi \equiv \pi(\Delta, \mathcal{A})$ denote the policy defined in Algorithm 1, and let $\mathbf{w} \equiv \{w_i \,:\, i \in \mathcal{N}\}$ be the profit vector. We fix $T \geq 2$ and consider arbitrary preferences $F^{(\mathbb{N})} \in \mathcal{F}$, where $\mathcal{F}$ is a class of preferences characterized by its magnitude $\mathcal{M}(\mathcal{F},T)$. Let $\Delta \in [T]$, and define $\tilde{T}-1 := \lceil T/\Delta \rceil - 1$ as the number of time sub-segments $\mathcal{F}_1,\ldots,\mathcal{F}_{\tilde{T}-1}$, each of size Δ (except possibly the last sub-segment, which may be smaller). We decompose the regret obtained by policy π into two components, denoted by $\mathcal{R}_1$ and $\mathcal{R}_2$, as follows:

$$J^*(F^{(\mathbb{N})},T) - J^\pi(F^{(\mathbb{N})},T) = \sum_{t=1}^{T}\big(r(S^*(F^t),F^t) - \mathbb{E}_{\mathbb{P}^\pi_{F^{(\mathbb{N})}}}\left[w_{i_t}\right]\big) \equiv \mathcal{R}_1 + \mathcal{R}_2,$$

where $\mathcal{R}_1$ and $\mathcal{R}_2$ are defined by:

$$\mathcal{R}_1 := \sum_{j=1}^{\tilde{T}-1}\Big(\sum_{t\in\mathcal{F}_j} r(S^*(F^t),F^t) - \max_{S\in\mathcal{S}}\sum_{t\in\mathcal{F}_j} r(S,F^t)\Big),$$

$$\mathcal{R}_2 := \sum_{j=1}^{\tilde{T}-1}\max_{S\in\mathcal{S}}\Big\{\sum_{t\in\mathcal{F}_j} r(S,F^t)\Big\} - \mathbb{E}_{\mathbb{P}^\pi_{F^{(\mathbb{N})}}}\Big[\sum_{t=1}^{T} w_{i_t}\Big].$$

The term $\mathcal{R}_1$ corresponds to the revenue loss incurred by replacing the fully informed oracle with a semi-oracle that knows the preferences for a time sub-segment but can only implement a single assortment for that sub-segment. Then, the term $\mathcal{R}_2$ captures the additional regret arising from employing policy π instead of the semi-oracle.

Step 1 (Upper bound for $\mathcal{R}_1$). We begin by deriving an upper bound for the first regret component $\mathcal{R}_1$. For any fixed $j \in [\tilde{T}-1]$, we define:

$$M_j = \sum_{t\in\mathcal{F}_j}\max\big\{\mathcal{K}^t(S) \,:\, S\in\mathcal{S}\big\},$$

as the cumulative preferences variation within sub-segment $\mathcal{F}_j$. Thus, by construction, if we sum over all sub-segment indices j, then we obtain $\sum_{j=1}^{\tilde{T}-1} M_j \leq \mathcal{M}(\mathcal{F},T)$.

Let $t_j \in \operatorname{argmin}\{p_0(S^*(F^t), F^t) : t \in \mathcal{F}_j\}$. We establish the following chain of inequalities:

$$\begin{aligned}
\sum_{t\in\mathcal{F}_j} r(S^*(F^t), F^t) - \max_{S\in\mathcal{S}} \sum_{t\in\mathcal{F}_j} r(S, F^t) &\leq \sum_{t\in\mathcal{F}_j} r(S^*(F^t), F^t) - \sum_{t\in\mathcal{F}_j} r(S^*(F^{t_j}), F^t) \\
&= \sum_{t\in\mathcal{F}_j} \sum_{i\in\mathcal{N}} w_i \big(p_i(S^*(F^t), F^t) - p_i(S^*(F^{t_j}), F^t)\big) \\
&\leq \|\mathbf{w}\|_1 \sum_{t\in\mathcal{F}_j} \big(\sum_{i\in\mathcal{N}} p_i(S^*(F^t), F^t) - \sum_{i\in\mathcal{N}} p_i(S^*(F^{t_j}), F^t)\big) \\
&= \|\mathbf{w}\|_1 \sum_{t\in\mathcal{F}_j} \big(p_0(S^*(F^{t_j}), F^t) - p_0(S^*(F^t), F^t)\big) \\
&\leq \|\mathbf{w}\|_1 \Delta \cdot \max_{t\in\mathcal{F}_j} \big\{p_0(S^*(F^{t_j}), F^t) - p_0(S^*(F^t), F^t)\big\}.
\end{aligned}$$

We proceed to prove that:

$$\max_{t\in\mathcal{F}_j} \big\{p_0(S^*(F^{t_j}), F^t) - p_0(S^*(F^t), F^t)\big\} \leq \sqrt{M_j/2},$$

via contradiction. Suppose there exists $t_0 \in \mathcal{F}_j$ such that:

$$p_0(S^*(F^{t_j}), F^{t_0}) - p_0(S^*(F^{t_0}), F^{t_0}) > \sqrt{M_j/2}.$$

As a consequence, we derive the following sequence of inequalities:

$$\begin{aligned}
\sqrt{M_j/2} &< p_0(S^*(F^{t_j}), F^{t_0}) - p_0(S^*(F^{t_0}), F^{t_0}) \\
&= p_0(S^*(F^{t_j}), F^{t_0}) - p_0(S^*(F^{t_j}), F^{t_j}) + p_0(S^*(F^{t_j}), F^{t_j}) - p_0(S^*(F^{t_0}), F^{t_0}) \\
&\overset{(a)}{\leq} p_0(S^*(F^{t_j}), F^{t_0}) - p_0(S^*(F^{t_j}), F^{t_j}) \\
&\overset{(b)}{\leq} \sum_{t=1}^{|\mathcal{F}_j|-1} \|p(S^*(F^{t_j}), F^{t+1}) - p(S^*(F^{t_j}), F^t)\|_\infty \\
&\overset{(c)}{\leq} \sum_{t=1}^{|\mathcal{F}_j|-1} \big(\frac{1}{2}\mathcal{K}(p(S^*(F^{t_j}), F^{t+1}), p\left(S^*(F^{t_j}), F^t\right)\,\big)^{\frac{1}{2}} \\
&\leq \big(\frac{1}{2}\sum_{t=1}^{|\mathcal{F}_j|-1} \mathcal{K}^t(S^*(F^{t_j}))\big)^{\frac{1}{2}} \leq \sqrt{M_j/2},
\end{aligned}$$

where (a) follows from the definition of t_j and (b) by the triangle inequality. Moreover, F^t refers to the t-th element of $\mathcal{F}_j$. Additionally, (c) follows from Pinkster's inequality (Tsybakov 2003). Therefore, we do have a contradiction as $\sqrt{M_j/2} < \sqrt{M_j/2}$.

Therefore, we conclude that:

$$\mathcal{R}_1 = \sum_{j=1}^{\tilde{T}-1} \big(\sum_{t\in\mathcal{F}_j} r(S^*(F^t), F^t) - \max_{S\in\mathcal{S}} \sum_{t\in\mathcal{F}_j} r(S, F^t)\big)$$

$$\leq \sum_{j=1}^{\tilde{T}-1} \|\mathbf{w}\|_1 \Delta \sqrt{M_j/2} \overset{(a)}{\leq} \frac{1}{\sqrt{2}} \Delta \sqrt{\tilde{T}-1} \|\mathbf{w}\|_1 \big(\sum_{j=1}^{\tilde{T}-1} M_j \big)^{\frac{1}{2}} \leq \frac{1}{\sqrt{2}} \Delta \sqrt{T/\Delta} \|\mathbf{w}\|_1 \cdot \sqrt{\mathcal{M}(\mathcal{F}, T)},$$

where (a) follows from the Jensen's inequality.

Step 2 (Upper bound for $\mathcal{R}_2$). We now establish a bound for the second regret component $\mathcal{R}_2$. For any fixed $j \in [\tilde{T}-1]$, let $\psi_{t,j}$ denote the assortment policy induced by $\mathcal{A}$ on $\mathcal{F}_j$ for $t \in \mathcal{F}_j$. Let $S \in \mathcal{S}$ fixed arbitrarily. For each sub-segment j, we identify $t_j \equiv t_j(S) \in \mathcal{F}_j$ such that:

$$\sum_{t \in \mathcal{F}_j} r(S, F^t) \leq \sum_{t \in \mathcal{F}_j} r(S, F^{t_j}).$$

Then, we have that:

$$\begin{aligned}
\sum_{t \in \mathcal{F}_j} r(S, F^t) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^t) &= \sum_{t \in \mathcal{F}_j} r(S, F^t) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^{t_j}) + \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^{t_j}) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^t) \\
&\leq \sum_{t \in \mathcal{F}_j} r(S, F^{t_j}) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^{t_j}) + \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^{t_j}) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^t) \\
&\overset{(a)}{=} \sum_{t \in \mathcal{F}_j} r(S, F^{t_j}) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^{t_j}) \\
&\quad + \sum_{t \in \mathcal{F}_j} \sum_{i \in [N]} w_i \big(p_i(\psi_{j,t}, F^{t_j}) - p_i(\psi_{j,t}, F^t) \big),
\end{aligned}$$

where (a) follows from the definition of the expected revenue.

Since $|\mathcal{S}|$ is finite, we have that:

$$\begin{aligned}
\sum_{t \in \mathcal{F}_j} r(S, F^{t_j}) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^{t_j}) &\overset{(a)}{\leq} \max_{S \in \mathcal{S}} \Big\{ \sum_{t \in \mathcal{F}_j} r(S, F^{t_j}) - \sum_{t \in \mathcal{F}_j} r(\psi_{j,t}, F^{t_j}) \Big\} \\
&\leq \sup_{F^{(\mathbb{N})} \in \mathcal{F}_S} \Big\{ \max_{S \in \mathcal{S}} \Big\{ \sum_{t \in \mathcal{F}_j} \big(r(S, F_S) - r(\psi_{j,t}, F_S) \big) \Big\} \;:\; F^t = F_S \; \forall \, t \in \mathbb{N} \Big\} \\
&= \mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_S, \Delta\big),
\end{aligned}$$

where (a) follows as the maximum is well-defined (since $t_j \equiv t_j(S)$ is defined for each $S \in \mathcal{S}$). Moreover, $\mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_S, \Delta\big)$ represents the minimax regret of policy $\mathcal{A}$ when preferences remain static.

Next, we derive the following sequence of inequalities:

$$\begin{aligned}
\sum_{t \in \mathcal{F}_j} \sum_{i \in [N]} w_i \big(p_i(\psi_{j,t}, F^{t_j}) - p_i(\psi_{j,t}, F^t) \big) &= \sum_{t \in \mathcal{F}_j} \|\mathbf{w}\|_\infty N \|p(\psi_{j,t}, F^{t_j}) - p(\psi_{j,t}, F^t)\|_\infty \\
&\overset{(a)}{\leq} \|\mathbf{w}\|_\infty N \sum_{t \in \mathcal{F}_j} \Big(\sum_{u=2}^{|\mathcal{F}_j|} \|p(\psi_{j,t}, F^u) - p(\psi_{j,t}, F^{u-1})\|_\infty \Big) \\
&\overset{(b)}{\leq} \|\mathbf{w}\|_\infty N \sum_{t \in \mathcal{F}_j} \Big(\sum_{u=2}^{|\mathcal{F}_j|} \sqrt{\frac{1}{2} \mathcal{K}^u(\psi_{j,t})} \Big)
\end{aligned}$$

$$\overset{(c)}{\leq} \|\mathbf{w}\|_\infty N \sum_{t\in\mathcal{F}_j} \big(\frac{1}{2}\sum_{u=2}^{|\mathcal{F}_j|} \mathcal{K}^u(\psi_{j,t})\big)^{1/2} \leq \|\mathbf{w}\|_1 N\Delta\sqrt{M_j/2},$$

where we denote by F^u the u-th element of $\mathcal{F}_j$. Moreover, (a) follows by the triangle inequality and (b) follows from the Pinsker's inequality (Tsybakov 2003). Then, (c) follows by Jensen's inequality.

Therefore, by summing over $j \in [\tilde{T}-1]$, and given that $\tilde{T}-1 = \lceil T/\Delta\rceil - 1$, we have that:

$$\begin{aligned}\sum_{j=1}^{\tilde{T}-1}\big(\sum_{t\in\mathcal{F}_j} r(S,F^t) - \sum_{t\in\mathcal{F}_j} r(\psi_{j,t},F^t)\big) &\overset{(a)}{\leq} \big(\tilde{T}-1\big)\mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_S,\Delta\big) + \frac{1}{\sqrt{2}}\Delta N\|\mathbf{w}\|_1\big(\sum_{j=1}^{\tilde{T}-1} M_j\big)^{\frac{1}{2}} \\ &= \big(\tilde{T}-1\big)\mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_S,\Delta\big) + \frac{1}{\sqrt{2}}\Delta N\|\mathbf{w}\|_1 \cdot \sqrt{\mathcal{M}(\mathcal{F},T)},\end{aligned}$$

where (a) follows from Jensen's inequality.

Specifically, we obtain the following upper bound on $\mathcal{R}_2$:

$$\sum_{j=1}^{\tilde{T}-1} \max_{S\in\mathcal{S}}\big\{\sum_{t\in\mathcal{F}_j}\big(r(S,F^t) - \mathbb{E}_{\mathbb{P}^\pi_{F^{(\mathbb{N})}}}[w_{i_t}]\big)\big\} \leq \lceil T/\Delta\rceil\, \mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_S,\Delta\big) + \frac{1}{\sqrt{2}}\Delta N\|\mathbf{w}\|_1\sqrt{\mathcal{M}(\mathcal{F},T)}.$$

Step 3 (Synthesis). Combining the bounds derived in Steps 1 and 2, we obtain:

$$J^*(F^{(\mathbb{N})},T) - J^\pi(F^{(\mathbb{N})},T) \leq \lceil T/\Delta\rceil\, \mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_S,\Delta\big) + \frac{1}{\sqrt{2}}\Delta\sqrt{T/\Delta}\|\mathbf{w}\|_1\big(N+1\big)\cdot\sqrt{\mathcal{M}(\mathcal{F},T)}.$$

Importantly, the right-hand side of the previous inequality does not depend on the customers' preferences $F^{(\mathbb{N})}$. Hence, by taking the supremum from both side of the inequality over $\mathcal{F}$, we obtain the desired result and conclude the proof. ■

E.C.2 Proofs for Section 5

First, in Section E.C.2.1, we consider scenarios in which the change cannot be detected passively. We derive the corresponding lower bound on achievable performances and an upper bound on the regret achieved by Algorithm 2 when information is available on the change's size. Then, in Section E.C.2.2, we focus on cases where the change is detectable by monitoring the sales from the pre-change optimal assortment. We derive both a lower bound on the achievable performance and an upper bound for the regret achieved by Algorithm 3. To streamline our discussion, some relevant notations, including the definitions of *minimum optimality gap* γ and *maximum revenue separation* δ as well as technical lemmas used within the proofs are relegated to Section E.C.4.

For $T \geq 2$ and $\Delta \in [T]$, we define $\tilde{T}-1 := \lceil T/\Delta \rceil - 1$ as the number of sub-segments $\mathcal{F}_1, \dots, \mathcal{F}_{\tilde{T}-1}$, each of size Δ. We also refer to these sub-segments interchangeably as "time segments" or "customer segments." Let $\ell_1 := 1$, and for $j \geq 2$, define $\ell_j := 1 + (j-1)\Delta$, with $\ell_{\tilde{T}} := T$. Throughout this section, we assume that F^1 and F^τ correspond to the pre- and post-change preferences, respectively. Specifically, preferences $F^{(\mathbb{N})} \equiv \left(F^t : t \in \mathbb{N}\right)$ are defined by $F^t := F^1$ for $t < \tau$ and $F^t := F^\tau$ for $t \geq \tau$, for some $\tau \in \mathbb{N}$, and satisfy $F^{(\mathbb{N})} \in \mathcal{F}_A$. We say that preferences $F^{(\mathbb{N})}$ are induced by F^1 and F^τ. We assume that both the post-change preferences F^τ and the change time τ are unknown to the retailer. Also, we define $\mathbb{P}^\pi_{\ell_j}$ as the distribution over customers purchase decisions across the T periods, conditional on the policy $\pi \in \mathcal{P}$ and the change occurring at time $\tau = \ell_j$.

E.C.2.1 Proofs for Section 5.2

We present the proofs for the setting in which the post-change preferences are unknown to the retailer and the change cannot be detected using information available solely from the pre-change optimal assortment. Specifically, we establish a lower bound on the regret that any admissible policy must incur, as stated in Proposition 1. We then derive an upper bound on the regret achieved by the robust planning strategy described in Algorithm 2, as formalized in Proposition 2.

Proof of Proposition 1. The proof closely follows the argument the proof of Proposition 5, which considers the case where the post-change preferences are known; refer to (E.C.3.1) for the complete proof. When the post-change preferences are unknown, no policy can achieve a regret smaller than the bound established in the known-setting case. The constant in that lower bound can be expressed in terms of both γ and ϕ, as specified in the definition of $\tilde{\mathcal{F}}_U$. Since the arguments remain unchanged, we omit the detailed derivation for brevity. ■

Next, we derive an upper bound on the regret achieved by Algorithm 2.

Proof of Proposition 2. Let $T \geq 2$, $\kappa > 0$, F^1 be such that $F^{(\mathbb{N})} \in \mathcal{F}_A$ and $\mathcal{A} \in \mathcal{P}$ be defined as in the proposition statement. Then, we denote by $F^{(\mathbb{N})} \in \tilde{\mathcal{F}}_U(F^1)$ customers' preferences for which the change cannot be detected based on the information available from the pre-change optimal assortment $S^*(F^1)$. Moreover, F^τ represents the post-change preferences, which remains unknown to the retailer. Then, we fix $\Delta_o = \sqrt{T}/\kappa^2$ and $\Delta_e = 4(\log T)/\kappa^2$.

Next, we segment the selling time horizon $[T]$ into sub-segments of size $\Delta_o + |\mathcal{E}|\Delta_e$. Specifically, we define $\ell_0 = 1$ and $\ell_j = \ell_{j-1} + \Delta_o + |\mathcal{E}|\Delta_e$, for $j \in [\tilde{T} - 1]$, where $\tilde{T} - 1 := \lceil T/(\Delta_o + |\mathcal{E}|\Delta_e) \rceil - 1$. Moreover, we introduce j^* as the index that satisfies $\ell_{j^*} < \tau \leq \ell_{j^*+1}$, where τ is the time-period at which the change happens. We denote by $\pi \equiv \pi(\kappa, F^1, \mathcal{E}, \mathcal{A})$ the policy defined in Algorithm 2.

Then, given some sub-segment index $j \in [\tilde{T} - 1]$, we define:

$$\hat{\Lambda}_{\ell_j} := \mathbf{1}\big(\max \big\{ \|p(S, \hat{F}(S)) - p(S, F^1)\|_\infty \, : \, S \in \mathcal{E} \big\} > \kappa/2 \big),$$

where $\hat{F}(S)$ corresponds to the empirical distribution of the purchase decisions conditional on the assortment $S \in \mathcal{E}$. Next, we introduce $\hat{k}$ as the stopping rule that is used within policy π. Specifically, given $j \in [\tilde{T} - 1]$, the stopping rule is formally defined by:

$$\hat{k}_{\ell_j}(\mathcal{H}_{\ell_j - 1}) := \ell_{j+1} \hat{\Lambda}_{\ell_j}.$$

Step 1. To begin, we define the assortment strategy that is induced by π as ψ_t, for $t \in [T]$. In particular, we omit the dependence of the policy on the filtration $(\mathcal{H}_t)_{t=0}^T$ to simplify the notations. Then, we derive the following upper bound for the difference between the expected revenue obtained by the oracle and the expected revenue obtained with π:

$$\begin{aligned}
J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) &= \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\tau-1} \big(r(S^*(F^1), F^1) - r(\psi_t, F^1) \big) \big] \\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\tau}^{T} \big(r(S^*(F^\tau), F^\tau) - r(\psi_t, F^\tau) \big) \big] \\
&\overset{(a)}{\leq} \delta \cdot \big(\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\tau-1} \mathbf{1}\big(\psi_t \neq S^*(F^1)\big) \big] + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\tau}^{T} \big(r(S^*(F^\tau), F^\tau) - r(\psi_t, F^\tau) \big] \big) \\
&\overset{(b)}{\leq} \delta \cdot \big(\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau - \hat{k})^+ \big] + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k} - \tau)^+ \big] + \sum_{S \in \mathcal{E}} \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\hat{k}} \mathbf{1}\big(\psi_t = S\big) \big] \big) \\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\max\{\hat{k}, \tau\}}^{T} \big(r(S^*(F^\tau), F^\tau) - r(\psi_t, F^\tau) \big) \big].
\end{aligned}$$

Note that, (a) follows from the definition of δ; recall (E.C.4). Moreover, there are two possible cases for implementing an assortment $S \neq S^*(F^1)$ before the change occurs. The first one corresponds to the case where the algorithm falsely detects a change earlier than τ and subsequently runs $\mathcal{A}$ to learn the new customers' preferences. The second one arises during an exploration batch, which requires implementing assortments in $\mathcal{E}$. Thus, (b) follows from these two observations, combined with the fact that implementing an assortment other than $S^*(F^\tau)$ after time period τ occurs either because the change has not yet been detected or because $\mathcal{A}$ is being executed to learn the new customers' preferences.

Importantly, the last part of the upper bound corresponds to the regret achieved by the policy $\mathcal{A}$ when customers' preferences are static. Formally, the following inequality holds:

$$\mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=\max\{\hat{k},\tau\}}^{T}\big(r\big(S^*(F^\tau),F^\tau\big)-r\big(\psi_t,F^\tau\big)\big)\Big] \leq \mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=\tau}^{T}\big(r\big(S^*(F^\tau),F^\tau\big)-r\big(\psi_t,F^\tau\big)\big)\Big]$$
$$\leq \mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_b(F^1),T\big),$$

where the last inequality follows from the definition of regret of the policy $\mathcal{A}$ for static preferences.

Step 2. We derive an upper bound for the probability of the Type I error, which we denote by $q_{f,j}$, for the test $\hat{\Lambda}_{\ell_j}$ from the sub-segment $j \in [\tilde{T}-1]$. Given $j \in [\tilde{T}-1]$, we define the purchase decisions within that segment by $Z^{\ell_j},\ldots,Z^{\ell_{j+1}-1}$. Then, we consider the following hypothesis test:

$$H_{0,j} : Z^{\ell_j},\ldots,Z^{\ell_{j+1}-1} \;\sim\; F^1$$
$$H_{1,j} : Z^{\ell_j},\ldots,Z^{\ell_{j+1}-1} \;\sim\; F \neq F^1.$$

Next, we derive an upper bound for $q_{f,j}$. We leverage from the multivariate Dvoretzky-Kiefer-Wolfowitz (shortly DKW) inequality by Naaman (2021). Specifically:

$$q_{f,j} = \mathbb{P}\big[\hat{\Lambda}_{\ell_j}=1 \mid H_{0,j}\big] \leq \mathbb{P}\big[\max\big\{\|p(S,\hat{F}(S))-p(S,F^1)\|_\infty \,:\, S\in\mathcal{E}\big\} > \kappa/2\big\} \mid H_{0,j}\big]$$
$$\overset{(a)}{\leq} K(\Delta_e+1)\exp\big(-\Delta_e\kappa^2/2\big) = K(4\kappa^{-2}(\log T)+1)T^{-2},$$

where (a) follows from the DKW inequality.

Step 3. Next, we derive an upper bound for the expression $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big]$. To proceed, we first derive the following sequence of inequalities:

$$\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big] = \sum_{u=1}^{\tau-1}\mathbb{P}^\pi_\tau\big[(\tau-\hat{k})^+ \geq u\big] = \sum_{u=1}^{\tau-1}\mathbb{P}^\pi_\tau\big[\hat{k}\leq \tau-u\big]$$
$$\overset{(a)}{\leq} \sum_{j=1}^{j^*}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\mathbb{P}^\pi_\tau\Big[\bigcup_{m=1}^{j}\{\hat{k}=\ell_m+1\}\Big]$$

$$\leq \sum_{j=1}^{j^*} \sum_{i=\ell_j}^{\ell_{j+1}-1} \sum_{m=1}^{j} \mathbb{P}_\tau^\pi\big[\hat{k} = \ell_m + 1\big]$$

$$\overset{(b)}{\leq} \sum_{j=1}^{j^*-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \sum_{m=1}^{j} q_{f,j} = \sum_{j=1}^{j^*-1} j|\mathcal{E}|\Delta_e q_{f,j},$$

where (a) follows from that $\tau \leq \ell_{j^*+1}$ and the definition of our stopping-time random variable. Then, (b) follows from the definition of the Type I error and that $\ell_{j^*} < \tau \leq \ell_{j^*+1}$.

Recall that each exploration batch is of size $\Delta_e = 4(\log T)/\kappa^2$. Therefore, we obtain the following upper bound for the expression $\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\tau - \hat{k})^+\big]$:

$$\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\tau - \hat{k})^+\big] \leq |\mathcal{E}|\Delta_e \frac{j^*(j^*-1)}{2} \max\{q_{f,j} \,:\, j \in [\tilde{T}-1]\}$$

$$\overset{(a)}{\leq} \frac{T^2}{2|\mathcal{E}|\Delta_e} K(4\kappa^{-2}(\log T) + 1)T^{-2} = K(\frac{1}{2|\mathcal{E}|} + \frac{\kappa^2}{8|\mathcal{E}|\log 2}),$$

where (a) follows from Step 2, in which a bound for $q_{f,j}$ is derived, and from the fact that $j^* \leq \tilde{T}-1$, implying $j^*(j^*-1) \leq (\tilde{T}-1)(\tilde{T}-2)$, as well as from the definition of Δ_e.

Step 4. In what follows, we provide an upper bound for the expression $\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k} - \tau)^+\big]$. To begin, let $q_{d,j}$ denote the probability of a Type II error for the hypothesis test defined above within the sub-segment $j \in [\tilde{T}-1]$. Then, the following sequence of inequalities holds:

$$\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k} - \tau)^+\big] = \sum_{u=0}^{T-\tau+1} \mathbb{P}_\tau^\pi\big[\hat{k} \geq \tau + u\big] \leq \sum_{j=j^*}^{\tilde{T}-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \mathbb{P}_\tau^\pi\Big[\bigcap_{m=j^*}^{j} \big\{\hat{k} \neq \ell_m + 1\big\}\Big]$$

$$\overset{(a)}{\leq} \sum_{j=j^*+2}^{\tilde{T}-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \mathbb{P}_\tau^\pi\Big[\bigcap_{m=j^*+2}^{j} \big\{\hat{k} \neq \ell_m + 1\big\}\Big] + 2|\mathcal{E}|\Delta_e$$

$$\overset{(b)}{=} |\mathcal{E}|\Delta_e\big(2 + \sum_{j=j^*+2}^{\tilde{T}-1} (q_{d,j^*+2})^{j-j^*-1}\big)$$

$$= |\mathcal{E}|\Delta_e\big(2 + q_{d,j^*+2}\frac{1-(q_{d,j^*+2})^{\tilde{T}-j^*}}{1-q_{d,j^*+2}}\big)$$

$$\leq \frac{2|\mathcal{E}|\Delta_e}{1-q_{d,j^*+2}} = \frac{8|\mathcal{E}|\log(T)/\kappa^2}{1-q_{d,j^*+2}},$$

where (a) follows from the observation that the change could occur anywhere within the sub-segment $\{\ell_{j^*}, \ldots, \ell_{j^*+1}-1\}$. Additionally, (b) holds because $q_{d,j} = q_{d,j^*+2}$ for all $j \in \{j^*+2, \ldots, \tilde{T}\}$, since the probability of a Type II error depends only on the occurrence of the change.

Step 5. Next, we derive an upper bound for the probability of the Type II error induced by the statistical test $\hat{\Lambda}_{\ell_j}$. To proceed, we assume that $\hat{\Lambda}_{\ell_j} = 0$. Hence, the following inequality holds:

$$\max\big\{\|p(S, \hat{F}(S)) - p(S, F^1)\|_\infty \,:\, S \in \mathcal{E}\big\} \leq \kappa/2.$$

Consequently, the following inequality is guaranteed to hold for all feasible assortments $S \in \mathcal{E}$:

$$\|\frac{1}{\Delta_e}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t(S) - p\left(S, F^1\right)\|_\infty \equiv \|p(S, \hat{F}(S)) - p\left(S, F^1\right)\|_\infty \leq \kappa/2,$$

where $Z^t(S)$ represents the purchase decision when assortment S is offered at time t, and a vector with only zero elements otherwise. Then, we derive the following sequence of inequalities, which is guaranteed to hold for all $S \in \mathcal{S}$:

$$\begin{aligned}
0 &< \|p\left(S, F^1\right) - p\left(S, F^\tau\right)\|_\infty \\
&\leq \|p\left(S, F^1\right) - \frac{1}{\Delta_e}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t(S)\|_\infty + \|\frac{1}{\Delta_e}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t(S) - p\left(S, F^\tau\right)\|_\infty \\
&\leq \kappa/2 + \|\frac{1}{\Delta_e}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t(S) - p\left(S, F^\tau\right)\|_\infty.
\end{aligned}$$

To proceed, we use the multivariate DKW inequality to derive an upper bound for the probability of the Type II error for the sub-segment j^*. Specifically:

$$\begin{aligned}
q_{d,j^*} &\leq \mathbb{P}\big[\hat{\Lambda}_{\ell_j} = 0 \mid H_{1,j^*}\big] \\
&\leq \mathbb{P}\big[\|\frac{1}{\Delta_e}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}} Z^t(S) - p(S, F^\tau)\|_\infty > \|p\left(S, F^1\right) - p\left(S, F^\tau\right)\|_\infty - \frac{\kappa}{2} \mid H_{1,j^*}\big] \\
&\leq K(\Delta_e + 1)\exp\big(-2\Delta_e(\|p(S, F^1) - p(S, F^\tau)\|_\infty - \frac{\kappa}{2})^2\big).
\end{aligned}$$

Hence, the following inequalities are guaranteed to hold:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k} - \tau)^+\big] &\leq 8|\mathcal{E}|\Delta_e\big(1 - K(\Delta_e + 1)\exp(-2\Delta_e(\|p(S, F^1) - p(S, F^\tau)\|_\infty - \frac{\kappa}{2})^2)\big)^{-1} \\
&\overset{(a)}{\leq} 8|\mathcal{E}|\Delta_e\big(1 - K(\Delta_e + 1)\exp(-2\Delta_e(\sqrt{\kappa/2} - \kappa/2)^2)\big)^{-1} \overset{(b)}{\leq} 8|\mathcal{E}|\Delta_e,
\end{aligned}$$

where (a) follows from the definition of κ together with the Pinsker's inequality (Tsybakov 2003). Moreover, (b) holds as long as $T \geq t(K, \kappa)$, where $t(K, \kappa)$ is the smallest sample size for which the probability of the Type II error q_{d,j^*} is bounded above by $1/2$.

Consequently, for any $T \geq t(K, \kappa)$, we obtain the following upper bound:

$$\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k} - \tau)^+\big] \leq \frac{8\kappa^{-2}|\mathcal{E}|\log T}{1 - q_{d,j^*+2}} \leq 16\kappa^{-2}|\mathcal{E}|\log T.$$

Step 6. Least but not last, we derive an upper bound for the term $\sum_{S\in\mathcal{E}} \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\hat{k}} \mathbf{1}\big(\psi_t = S\big)\big]$. Importantly, by the definition of the stopping time $\hat{k}$, we must have $\hat{k} \leq T$. Hence, we obtain the following sequence of inequalities, which holds for any $T \geq 2$:

$$\sum_{S\in\mathcal{E}} \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\hat{k}} \mathbf{1}\big(\psi_t = S\big)\big] \leq \sum_{S\in\mathcal{E}} (\tilde{T}-1)\Delta_e \leq \sum_{S\in\mathcal{E}} \frac{T}{\Delta_o + |\mathcal{E}|\Delta_e}\Delta_e$$

$$= |\mathcal{E}| \frac{4T(\log T)/(\kappa^2)}{\sqrt{T}/\kappa^2 + 4\,|\mathcal{E}|\,(\log T)/(\kappa^2)} \leq 4|\mathcal{E}|\sqrt{T}\log T.$$

Step 7. We conclude the proof by aggregating the upper bounds that are derived in the previous steps (specifically, in Steps 3, 4 and 6). Importantly, we assume that the time horizon is large enough, that is, $T \geq t(K,\kappa)$. Under this assumption, we obtain the following upper bound on the difference between the expected revenue obtained by the oracle and the one from policy π:

$$J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) \leq \delta \cdot \big(K(\frac{1}{2|\mathcal{E}|} + \frac{\kappa^2}{8|\mathcal{E}|\log 2}) + 16\kappa^{-2}|\mathcal{E}|\log T + 4|\mathcal{E}|\sqrt{T}\log T\big)$$
$$+ \mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_b(F^1), T\big).$$

Therefore, we obtain the following upper bound on the regret of our policy:

$$J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) \leq C_1 + C_2 \log T + 4\|\mathbf{w}\|_1 |\mathcal{E}|\sqrt{T}\log T + \mathcal{R}^{\mathcal{A}}\big(\mathcal{F}_b(F^1), T\big),$$

where we use that $\delta \leq \|\mathbf{w}\|_1$ and set:

$$C_1 \equiv C_1(K,\kappa,\mathcal{E},\Delta) := K\|\mathbf{w}\|_1 \cdot \big(\frac{1}{2|\mathcal{E}|} + \frac{\kappa^2}{8|\mathcal{E}|\log 2}\big), \quad \text{and} \quad C_2 \equiv C_2(\kappa,\mathcal{E}) := 16\|\mathbf{w}\|_1 \kappa^{-2}|\mathcal{E}|.$$

To conclude, if we take the supremum over all possible instances $F^{(\mathbb{N})} \in \tilde{\mathcal{F}}_U(F^1)$ of customers' preferences of the previous expression, then we obtain the desired result. ∎

E.C.2.2 Proofs for Section 5.3

To proceed, we present the proofs of the results obtained when the post-change preferences are unknown but the change can be distinguished based on the available information from $S^*(F^1)$. Specifically, we establish a lower bound on the regret that any admissible policy must achieve, as described in the proof of Proposition 3. Also, we derive an upper bound on the regret achieved by the robust planning strategy described in Algorithm 3 as outlined in proof of Proposition 4.

Proof of Proposition 3. The proof closely follows the one for Proposition 7 in the case where the post-change preferences are assumed to be known by the retailer. Importantly, if the post-change preferences are unknown, then no policy can achieve a regret lower than that in the known case. Also, the constant ϑ in the lower bound of Proposition 7 can be replaced by ε (see the proof of Proposition 7 for details) and the results then coincide. Therefore, we omit the proof for brevity. ∎

Next, we derive an upper bound on the regret achieved by Algorithm 3.

Proof of Proposition 4. Let $T \geq 2$, $\varepsilon > 0$, F^1 be such that $F^{(\mathbb{N})} \in \mathcal{F}_A$ and $\mathcal{A} \in \mathcal{P}$ be defined as in the proposition statement. Then, we denote by $F^{(\mathbb{N})} \in \mathcal{F}_D(F^1)$ the customers' preferences, where a change can be detected based on the information available from the pre-change optimal assortment $S^*(F^1)$. Moreover, F^τ represents the post-change preferences, which remain unknown to the retailer. Then, we fix $\Delta = C \log T$, where $C := 4\varepsilon^{-2}$, as specified in the policy.

Next, we define $\ell_0 = 1$ and $\ell_{j+1} = \ell_j + \Delta$, for $j \in \{0, \ldots, \tilde{T} - 1\}$, where $\tilde{T} := \lceil T/\Delta \rceil$. Moreover, we introduce j^* as the index which satisfies $\ell_{j^*} < \tau \leq \ell_{j^*+1}$, where τ is the time period at which the change happens. We denote by $\hat{k}$ the stopping rule that is used within policy π from Algorithm 3. Specifically, we have:

$$\hat{k}_{\ell_j}(\mathcal{H}_{\ell_j - 1}) := \ell_{j+1} \mathbf{1}\big(\|\hat{p} - p\left(S^*(F^1), F^1\right)\|_\infty > \varepsilon/2\big),$$

where $\hat{p}$ denotes the empirical purchase distribution conditional on $S^*(F^1)$.

Step 1. We introduce the assortment strategy determined by the policy π at time $t \in [T]$, and which we denote by ψ_t. For simplicity, we omit the explicit dependence of ψ_t on the filtration $(\mathcal{H}_t)_{t=0}^T$. We then derive the following inequalities to bound the difference between the oracle's expected revenue and the expected revenue achieved by the policy:

$$\begin{aligned}
J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) &= \mathbb{E}_{\mathbb{P}_\tau^\pi}\Big[\sum_{t=1}^{\tau-1} \big(r(S^*(F^1), F^1) - r(\psi_t, F^1)\big)\Big] \\
&\quad - \mathbb{E}_{\mathbb{P}_\tau^\pi}\Big[\sum_{t=\tau}^{T} \big(r(S^*(F^\tau), F^\tau) - r(\psi_t, F^\tau)\big)\Big] \\
&\overset{(a)}{\leq} \delta \cdot \mathbb{E}_{\mathbb{P}_\tau^\pi}\Big[\sum_{t=1}^{\tau-1} \mathbf{1}\left(\psi_t \neq S^*(F^1)\right)\Big] \\
&\quad + \mathbb{E}_{\mathbb{P}_\tau^\pi}\Big[\sum_{t=\tau}^{T} \big(r\left(S^*(F^\tau), F^\tau\right) - r(\psi_t, F^\tau)\big)\Big] \\
&= \delta \cdot \Big(\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\tau - \hat{k})^+\big] + \mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k} - \tau)^+\big]\Big) \\
&\quad + \mathbb{E}_{\mathbb{P}_\tau^\pi}\Big[\sum_{t=\max\{\hat{k}, \tau\}}^{T} \big(r(S^*(F^\tau), F^\tau) - r(\psi_t, F^\tau)\big)\Big],
\end{aligned}$$

where (a) follows from the definition of δ.

Hence, the regret can be decomposed into two distinct terms: the delay associated with our change detection approach and the regret that is driven by learning the new customers' preferences. In particular, the later can be bounded above as follows:

$$\mathbb{E}_{\mathbb{P}_\tau^\pi}\Big[\sum_{t=\max\{\hat{k}, \tau\}}^{T} \big(r\left(S^*(F^\tau), F^\tau\right) - r\left(\psi_t, F^\tau\right)\big)\Big] \leq \mathbb{E}_{\mathbb{P}_\tau^\pi}\Big[\sum_{t=\tau}^{T} \big(r\left(S^*(F^\tau), F^\tau\right) - r\left(\psi_t, F^\tau\right)\big)\Big],$$

which essentially captures the regret incurred by the dynamic assortment planning policy designed for static preferences (arising from the subroutine $\mathcal{A}$, which is invoked within our policy) over the selling time horizon from τ to T. This term quantifies the performance gap between the optimal policy and the choices made by the dynamic assortment planning policy designed for static preferences. Hence, it is naturally bounded above by $\mathcal{R}^{\mathcal{A}}(\mathcal{F}_b(F^1), T)$.

Step 2. We establish a bound for the probability of the Type I error in the hypothesis test that is conducted within the policy for each segment. First, we denote it by $q_{f,j}$. To proceed, we start by fixing an arbitrary index $j \in [\tilde{T}-1]$. Then, we assume that the following purchase decisions $Z^{\ell_j}, \ldots, Z^{\ell_{j+1}-1}$ are observed. The hypothesis test is thus formally defined as follows:

$$H_{0,j} : Z^{\ell_j}, \ldots, Z^{\ell_{j+1}-1} \sim F^1 \;\big|\; S^*(F^1),$$
$$H_{1,j} : Z^{\ell_j}, \ldots, Z^{\ell_{j+1}-1} \sim F \neq F^1 \;\big|\; S^*(F^1).$$

Next, define $\hat{\Lambda}_{\ell_j} := \mathbf{1}\big(\|\hat{p} - p(S^*(F^1), F^1)\|_\infty > \varepsilon/2\big)$, where $\hat{p}$ represents the empirical purchase distribution conditional on $S^*(F^1)$, as the statistical test for the above hypothesis test. By definition, the probability of the Type I error $q_{f,j}$ is given by $q_{f,j} := \mathbb{P}\big[\hat{\Lambda}_{\ell_j} = 1 \;\big|\; H_{0,j}\big]$. To bound this probability, we use that each assortment has at most K products, i.e., $\|S\|_1 \leq K$, for each assortment $S \in \mathcal{S}$ and then apply the multivariate DKW inequality (Naaman 2021). Specifically:

$$\mathbb{P}\big[\hat{\Lambda}_{\ell_j} = 1 \;\big|\; H_{0,j}\big] \leq K(\Delta + 1)\exp(-\frac{1}{2}\Delta\varepsilon^2).$$

Step 3. Next, we find an upper bound for the expression $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau - \hat{k})^+\big]$. We first fix some index $j \in [j^*]$. Then, we proceed by deriving the following sequence of inequalities:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau - \hat{k})^+\big] &= \sum_{u=1}^{\tau-1} \mathbb{P}^\pi_\tau\big[(\tau - \hat{k})^+ \geq u\big] = \sum_{u=1}^{\tau-1} \mathbb{P}^\pi_\tau\big[\hat{k} \leq \tau - u\big] \\
&\overset{(a)}{\leq} \sum_{j=1}^{j^*} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \mathbb{P}^\pi_\tau\Big[\bigcup_{m=1}^{j} \{\hat{k} = \ell_m + 1\}\Big] \\
&\leq \sum_{j=1}^{j^*} \sum_{i=\ell_j}^{\ell_{j+1}-1} \sum_{m=1}^{j} \mathbb{P}^\pi_\tau\big[\hat{k} = \ell_m + 1\big] \overset{(b)}{\leq} \sum_{j=1}^{j^*-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \sum_{m=1}^{j} q_{f,j} = \Delta \sum_{j=1}^{j^*-1} j q_{f,j},
\end{aligned}$$

where (a) follows from that $\tau \leq \ell_{j^*+1}$ and the definition of our stopping-time random variable. Then, (b) follows from the definition of the Type I error and that $\ell_{j^*} < \tau \leq \ell_{j^*+1}$.

Therefore, we have that:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau - \hat{k})^+\big] &\leq \Delta \frac{j^*\,(j^* - 1)}{2} q_{f,j} \\
&\leq K\Delta \frac{\tilde{T}(\tilde{T} - 1)}{2} \exp(-\frac{1}{2}\Delta\varepsilon^2)
\end{aligned}$$

$$\leq \Delta^{-1}T^2K(\Delta+1)\exp(-\frac{1}{2}\Delta\varepsilon^2) = K(1+\Delta^{-1}) \leq K(1+\frac{\varepsilon^2}{4\log(2)}).$$

Step 4. The next step provides an upper bound for the expression $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k}-\tau)^+\big]$. We denote by $q_{d,j}$ the probability of the Type II error of the test within the sub-segment j:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k}-\tau)^+\big] &= \sum_{u=0}^{T-\tau+1} \mathbb{P}^\pi_\tau\big[\hat{k} \geq \tau+u\big] \\
&\leq \sum_{j=j^*}^{\tilde{T}-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \mathbb{P}^\pi_\tau\big[\bigcap_{m=j^*}^{j} \{\hat{k} \neq \ell_m+1\}\big] \\
&\overset{(a)}{\leq} \sum_{j=j^*+2}^{\tilde{T}-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \mathbb{P}^\pi_\tau\big[\bigcap_{m=j^*+2}^{j} \{\hat{k} \neq i_m+1\}\big] + 2\Delta \\
&\overset{(b)}{=} \Delta\big(2+\sum_{j=j^*+2}^{\tilde{T}-1} (q_{d,j^*+2})^{j-j^*-1}\big) = \Delta\big(2+q_{d,j^*+2}\frac{1-(q_{d,j^*+2})^{\tilde{T}-j^*}}{1-q_{d,j^*+2}}\big) \leq \frac{2\Delta}{1-q_{d,j^*+2}},
\end{aligned}$$

where (a) follows from the fact that the change could occur anywhere within the customer segment $\{\ell_{j^*},\ldots,\ell_{j^*+1}-1\}$. Moreover, (b) holds because $q_{d,j} = q_{d,j^*+2}$ for all $j \in \{j^*+2,\ldots,\tilde{T}\}$. This equivalence arises from the fact that the probability of the Type II error depends solely on the occurrence of the change, rather than on its specific sub-segment after index j^*+2.

Step 5. In the following, we provide an upper bound for the Type II error at the sub-segment j^*. To proceed, we assume that the statistical test does not reject the null hypothesis within sub-segment j^*. That is, we assume that:

$$\|\frac{1}{\Delta}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t - p\left(S^*(F^1),F^1\right)\|_\infty \leq \frac{\varepsilon}{2}.$$

Therefore, the following sequence of inequalities holds:

$$\begin{aligned}
0 &< \|p\left(S^*(F^1),F^1\right) - p\left(S^*(F^1),F^\tau\right)\|_\infty \\
&\leq \|p\left(S^*(F^1),F^1\right) - \frac{1}{\Delta}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t\|_\infty + \|\frac{1}{\Delta}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t - p\left(S^*(F^1),F^\tau\right)\|_\infty \\
&\leq \frac{\varepsilon}{2} + \|\frac{1}{\Delta}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}-1} Z^t - p\left(S^*(F^1),F^\tau\right)\|_\infty.
\end{aligned}$$

Hence, by using the multivariate DKW inequality, we derive the following upper bound for the probability of the Type II error within the sub-segment j^*:

$$q_{d,j^*} \leq \mathbb{P}\big[\hat{\Lambda}_{\ell_{j^*}} = 0 \mid H_{1,j^*}\big]$$

$$\leq \mathbb{P}\big[\|\frac{1}{\Delta}\sum_{t=\ell_{j^*}}^{\ell_{j^*+1}} Z^t - p(S^*(F^1), F^\tau)\|_\infty > \|p\left(S^*(F^1), F^1\right) - p\left(S^*(F^1), F^\tau\right)\|_\infty - \frac{\varepsilon}{2} \mid H_{1,j^*}\big]$$

$$\leq K(\Delta+1)\exp\big(-2\Delta(\|p(S^*(F^1),F^1) - p(S^*(F^1),F^\tau)\|_\infty - \frac{\varepsilon}{2})^2\big).$$

Hence, the following inequalities hold:

$$\mathbb{E}_{\mathbb{P}^\pi_\tau}\left[(\hat{k}-\tau)^+\right] \leq 2\Delta\big(1 - K(\Delta+1)\exp(-2\Delta(\|p(S^*(F^1),F^1) - p(S^*(F^1),F^\tau)\|_\infty - \frac{\varepsilon}{2})^2)\big)^{-1}$$
$$\stackrel{(a)}{\leq} 2\Delta\big(1 - K(\Delta+1)\exp(-2\Delta(\sqrt{\varepsilon/2} - \varepsilon/2)^2)\big)^{-1} \stackrel{(b)}{\leq} 6\Delta,$$

where (a) follows from the definition of ε together with the Pinsker's inequality (Tsybakov 2003). Moreover, (b) holds as long as the time horizon is large enough, that is, $T \geq t(\varepsilon, K)$, where $t(\varepsilon, K)$ is the smallest integer which guarantees that $q_{d,j^*} < 1/2$.

Step 6. Finally, we aggregate all the bounds that we obtain within the previous steps (specifically, in Steps 3 and 5). Recall that $\Delta = C\log T$, where $C = 4\varepsilon^{-2}$. Thus, we derive the following upper bound on the difference in the expected revenue achieved by the oracle and our policy:

$$\begin{aligned} J^*(F^{(\mathbb{N})},T) - J^\pi(F^{(\mathbb{N})},T) &\leq \delta K(1+\frac{\varepsilon^2}{4\log 2}) + 2\delta\Delta + \mathcal{R}^{\mathcal{A}}(\mathcal{F}_b(F^1),T) \\ &= \delta K(1+\frac{\varepsilon^2}{4\log 2}) + 2C\log T\delta + \mathcal{R}^{\mathcal{A}}(\mathcal{F}_b(F^1),T) \\ &\leq \delta K(1+\frac{\varepsilon^2}{4\log 2}) + 8\varepsilon^{-2}\delta\log T + \mathcal{R}^{\mathcal{A}}(\mathcal{F}_b(F^1),T). \end{aligned}$$

Finally, we define $C_1 \equiv C_1(\varepsilon,\delta) := \delta K(1+\frac{\varepsilon^2}{4\log 2})$, and $C_2 \equiv C_2(\varepsilon,\delta) := 8\varepsilon^{-2}\delta$. Hence, by taking the supremum over all customers' preferences, we obtain the following bound for the regret:

$$\sup_{F^{(\mathbb{N})}\in\mathcal{F}_D(F^1)} \big\{J^*(F^{(\mathbb{N})},T) - J^\pi(F^{(\mathbb{N})},T)\big\} \leq C_1 + C_2\log T + \mathcal{R}^{\mathcal{A}}(\mathcal{F}_b(F^1),T),$$

for $T \geq t(\varepsilon, K)$, which, in turn, concludes the proof. ■

E.C.3 Proofs for Appendix A.3

This section contains the proofs for the results established in Appendix A.3, which addresses the case where the post-change preferences are known by the retailer. Our analysis proceeds in three stages. In Section E.C.3.1, we treat the case in which changes are passively undetectable. Next, in Section E.C.3.2, we consider the case with passively detectable changes. We provide both a lower bound on the achievable regret for any admissible policy and an upper bound on the regret attained by our policies. Finally, in Section E.C.3.3, we prove two lemmas that are used within the proofs.

E.C.3.1 Proofs for Appendix A.3.1

This section presents proofs for Proposition 5 and Proposition 6, about achievable performances and regret of Algorithm 4. To begin, we fix $T \geq 2$ and a sub-segment size $\Delta \in [T]$. Also, let $\tilde{T} := \lceil T/\Delta \rceil$ denote the number of time sub-segments. We define the sub-segment boundaries $\ell_1 := 1$, $\ell_j := 1+(j-1)\Delta$ for $j \geq 2$, and $\ell_{\tilde{T}} := T$. We assume throughout that F^1 and F^τ are such that that the preferences $F^{(\mathbb{N})}$ they induce belong to $\mathcal{F}_A$. For any policy $\pi \in \mathcal{P}$, let $\mathbb{P}^\pi_{\ell_j}$ denote the probability distribution of the customer purchase decisions over T periods when the change occurs at $\tau = \ell_j$.

The analysis proceeds through two essential lemmas that address distinct policy classes. First, Lemma 1 analyzes policies that do not sufficiently explore alternative assortments within at least one sub-segment. Then, Lemma 2 considers policies that are always guaranteed to sufficiently explore new product assortments. We quantify exploration intensity through the KL divergence between successive scenarios $\mathbb{P}^\pi_{\ell_{j+1}}$ and $\mathbb{P}^\pi_{\ell_j}$. From a high-level perspective, a small KL divergence indicates a minimal difference in customers' preferences across adjacent change points, suggesting limited exploration within the corresponding segment. Importantly, our approach relies on probabilistic arguments by Besbes and Zeevi (2011) and by Tsybakov (2003).

Lemma 1. *Assume that there exist an admissible policy $\pi \in \mathcal{P}$ and a constant $\beta > 0$ that satisfy:*

$$\min_{1 \leq j \leq \tilde{T}-1} \left\{ \mathcal{K}\big(\mathbb{P}^\pi_{\ell_{j+1}}, \mathbb{P}^\pi_{\ell_j}\big) \right\} \leq \beta.$$

Then, there exists a finite constant $C \equiv C(\gamma, \beta) > 0$ such that:

$$\sup_{F^{(\mathbb{N})} \in \tilde{\mathcal{F}}_U} \left\{ J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) \right\} \geq C\Delta.$$

Proof. We fix some preferences F^1 and F^τ and consider the induced preferences $F^{(\mathbb{N})} \in \tilde{\mathcal{F}}_U(F^1, F^\tau)$. Also, we assume that there exist an admissible policy $\pi \in \mathcal{P}$ and a constant $\beta > 0$ such that:

$$\min_{1 \leq j \leq \tilde{T}-1} \{\mathcal{K}(\mathbb{P}^\pi_{\ell_{j+1}}, \mathbb{P}^\pi_{\ell_j})\} \leq \beta.$$

Next, let $i_0 \in [\tilde{T}-1]$ be such that $\mathcal{K}(\mathbb{P}^{\pi}_{\ell_{i_0+1}}, \mathbb{P}^{\pi}_{\ell_{i_0}}) \leq \beta$. We consider the following two hypotheses:

$$H_0: \quad \tau \notin \{\ell_{i_0}, \ldots, \ell_{i_0+1} - 1\},$$
$$H_1: \quad \tau = \ell_{i_0}.$$

Under the probability measure $\mathbb{P}^{\pi}_{\ell_{i_0}}$, the distribution of the customer's purchase decisions undergoes a shift at ℓ_{i_0}, and changes from F^1 to F^{τ}. Conversely, under $\mathbb{P}^{\pi}_{\ell_{i_0+1}}$, no such shift occurs within the interval $\{\ell_{i_0}, \ldots, \ell_{i_0+1} - 1\}$.

Next, we define an arbitrary admissible decision rule:

$$\phi : \mathcal{S}^{\ell_{i_0+1}-1} \times \{0,1\}^{\ell_{i_0+1}-1} \to \{0,1\},$$

where $\phi = 0$ indicates that "no change," occurs before $\ell_{i_0+1} - 1$, which essentially implies that $\tau \notin \{\ell_{i_0}, \ldots, \ell_{i_0+1} - 1\}$, whereas $\phi = 1$ indicates that a change has occurred precisely at ℓ_{i_0}. Thus, ϕ maps the set of all possible assortments and the corresponding purchase decisions observed from customers 1 to $\ell_{i_0+1} - 1$ to $\{0,1\}$. According to Theorem 2.2 of Tsybakov 2003, we have:

$$\inf_{\phi} \max\{\mathbb{P}^{\pi}_{\ell_{i_0}}[\phi \neq 1], \mathbb{P}^{\pi}_{\ell_{i_0+1}}[\phi \neq 0]\} \geq \max\Big\{\frac{1}{4}\exp(-\beta), \frac{1-\sqrt{\beta/2}}{2}\Big\}.$$

In other words, the following inequality holds:

$$\inf_{\phi} \min\{\mathbb{P}^{\pi}_{\ell_{i_0}}[\phi = 0], \mathbb{P}^{\pi}_{\ell_{i_0+1}}[\phi = 1]\} \geq \max\Big\{\frac{1}{4}\exp(-\beta), \frac{1-\sqrt{\beta/2}}{2}\Big\}.$$

In addition, we define the following constant:

$$\tilde{C} := \frac{\gamma}{4} \max\Big\{\frac{1}{4}\exp(-\beta), \frac{1-\sqrt{\beta/2}}{2}\Big\},$$

where $\gamma > 0$ by definition.

Then, suppose, for the sake of contradiction, that the following inequality holds:

$$\sup_{k \in \{i_0, i_0+1\}} \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_k}}\left[J^*(F^{(\mathbb{N})}, T) - \mathcal{J}^{\pi}(F^{(\mathbb{N})}, T)\right] \leq \tilde{C}\Delta. \tag{E.C.32}$$

Next, consider the following decision rule:

$$\phi(\pi) = \begin{cases} 0 & \text{if } \sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1} \big(r(S^*(F^1), F^1) - r(\psi_t(\mathcal{H}_{t-1}), F^1)\big) \leq \gamma\Delta/2, \\ 1 & \text{if } \sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1} \big(r(S^*(F^1), F^1) - r(\psi_t(\mathcal{H}_{t-1}), F^1)\big) > \gamma\Delta/2, \end{cases}$$

where the decision rule ϕ implicitly depends on the observed realization of the purchase decisions through the filtration $(\mathcal{H}_t)_{t=0}^{\ell_{i_0+1}-1}$. We complete the proof in three steps by deriving upper bounds for both the Type I and II errors of our decision rule, and then combining them.

Step 1: We first establish an upper bound for the Type I error probability, the probability of incorrectly rejecting the null hypothesis by indicating a change when none has occurred before sub-segment i_0+1. The probability of the Type I error can be formally expressed as:

$$\begin{aligned}
\mathbb{P}^{\pi}_{\ell_{i_0+1}}\big[\phi=1\big] &= \mathbb{P}^{\pi}_{\ell_{i_0+1}}\Big[\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\Big(r\left(S^*(F^1),F^1\right)-r\left(\psi_t(\mathcal{H}_{t-1}),F^1\right)\Big)>\frac{\gamma}{2}\Delta\Big]\\
&\overset{(a)}{\leq} \frac{2}{\gamma\Delta}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{i_0+1}}}\Big[\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\Big(r\left(S^*(F^1),F^1\right)-r\left(\psi_t(\mathcal{H}_{t-1}),F^1\right)\Big)\Big]\\
&\overset{(b)}{\leq} \frac{2}{\gamma\Delta}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{i_0+1}}}\left[J^*(F^{(\mathbb{N})},T)-\mathcal{J}^{\pi}(F^{(\mathbb{N})},T)\right]\\
&\overset{(c)}{\leq} \frac{2\tilde{C}}{\gamma}=\frac{1}{2}\max\Big\{\frac{1}{4}\exp(-\beta),\frac{1-\sqrt{\beta/2}}{2}\Big\}
\end{aligned}$$

The derivations above employ three key steps: (a) applies Markov's inequality (Jacod and Protter 2012), (b) follows from the optimality of $S^*(F^1)$ and the definition of $\mathcal{J}^{\pi}(F^{(\mathbb{N})},T)$ as established in Lemma 5, and (c) uses the bound from equation (E.C.32).

Step 2: We now establish an upper bound for the probability of the Type II error, defined as the probability that our decision rule fails to detect a change when one has actually occurred. To proceed, let assume that $\phi=0$. Under these conditions, the following inequality is satisfied:

$$\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\Big(r\left(S^*(F^1),F^1\right)-r\left(\psi_t(\mathcal{H}_{t-1}),F^1\right)\Big)\leq\gamma\Delta/2.$$

In particular, the following inequality also holds:

$$\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\Big(r\left(S^*(F^1),F^1\right)-r\left(\psi_t(\mathcal{H}_{t-1}),F^1\right)\Big)\mathbf{1}\left(\|S^*(F^1)-\psi_t(\mathcal{H}_{t-1})\|_1\geq 1\right)\leq\gamma\Delta/2,$$

and, we obtain the following inequality:

$$\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\mathbf{1}\left(\|S^*(F^1)-\psi_t(\mathcal{H}_{t-1})\|_1\geq 1\right)\leq\frac{\gamma\Delta}{2\gamma}=\Delta/2.$$

Therefore, we obtain the following sequence of inequalities:

$$\begin{aligned}
\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\Big(r\left(S^*(F^\tau),F^\tau\right)-r\left(\psi_t(\mathcal{H}_{t-1}),F^\tau\right)\Big) &\geq \gamma\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\mathbf{1}\left[\|S^*(F^1)-\psi_t(\mathcal{H}_{t-1})\|_1\leq 0\right]\\
&=\gamma\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\Big(1-\mathbf{1}\left[\|S^*(F^1)-\psi_t(\mathcal{H}_{t-1})\|_1\geq 1\right]\Big)\\
&=\gamma\Big(\Delta-\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\mathbf{1}\left[\|S^*(F^1)-\psi_t(\mathcal{H}_{t-1})\|_1\geq 1\right]\Big)\\
&=\gamma\big(\Delta-\Delta/2\big)=\gamma\Delta/2.
\end{aligned}$$

As a consequence, we derive the following upper bounds for the Type II errors of ϕ:

$$\mathbb{P}^{\pi}_{\ell_{i_0}}\big[\phi=0\big] \leq \mathbb{P}^{\pi}_{\ell_{i_0}}\Big[\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\big(r\left(S^*(F^\tau),F^\tau\right)-r\left(\psi_t(\mathcal{H}_{t-1}),F^\tau\right)\big) \geq \frac{\gamma\Delta}{2}\Big]$$

$$\stackrel{(a)}{\leq} \frac{2}{\gamma\Delta}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{i_0}}}\Big[\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\big(r\left(S^*(F^\tau),F^\tau\right)-r\left(\psi_t(\mathcal{H}_{t-1}),F^\tau\right)\big)\Big]$$

$$\stackrel{(b)}{\leq} \frac{2}{\gamma\Delta}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{i_0}}}\big[J^*(F^{(\mathbb{N})},T)-\mathcal{J}^{\pi}(F^{(\mathbb{N})},T)\big] \stackrel{(c)}{\leq} \frac{2}{\gamma\Delta}\tilde{C}\Delta = \frac{1}{2}\max\Big\{\frac{1}{4}\exp(-\beta),\frac{1-\sqrt{\beta/2}}{2}\Big\},$$

where (a) follows from Markov's inequality (Jacod and Protter 2012), while (b) follows from the optimality of $S^*(F^\tau)$ and Lemma 5. Finally the last inequality (c) follows from (E.C.32).

Step 3: Consequently, based on the results from both Step 1 and Step 2, we conclude that the infimum of the Type I and Type II errors is bounded above as follows:

$$\inf_{\phi}\min\big\{\mathbb{P}^{\pi}_{\ell_{i_0}}\left[\phi=0\right],\mathbb{P}^{\pi}_{\ell_{i_0+1}}\left[\phi=1\right]\big\} \leq \frac{1}{2}\max\Big\{\frac{1}{4}\exp(-\beta),\frac{1-\sqrt{\beta/2}}{2}\Big\},$$

which is a contradiction with (E.C.32). Hence, the following inequality must hold:

$$\sup_{k\in\{i_0,i_0+1\}}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_k}}\big[J^*(F^{(\mathbb{N})},T)-\mathcal{J}^{\pi}(F^{(\mathbb{N})},T)\big] > \tilde{C}\Delta.$$

Therefore, we conclude that:

$$\sup_{F^{(\mathbb{N})}\in\tilde{\mathcal{F}}_U}\big\{J^*(F^{(\mathbb{N})},T)-J^{\pi}(F^{(\mathbb{N})},T)\big\} > C\left(\gamma,\beta\right)\Delta,$$

where $C\left(\gamma,\beta\right)=\frac{\gamma}{4}\max\big\{\frac{1}{4}\exp(-\beta),\frac{1-\sqrt{\beta/2}}{2}\big\}$, where β is as defined in the proposition. ■

We proceed to establish a lower bound on the attainable regret for admissible policies that exhibit adequate exploration within each sub-segment. Formally:

Lemma 2. *Assume that there exist an admissible policy $\pi\in\mathcal{P}$ and a constant $\beta>0$ that satisfy:*

$$\min_{1\leq j\leq\tilde{T}-1}\big\{\mathcal{K}\big(\mathbb{P}^{\pi}_{\ell_{j+1}},\mathbb{P}^{\pi}_{\ell_j}\big)\big\} > \beta.$$

Then, there exists a finite constant $C\equiv C(\gamma,\vartheta,\beta)>0$ such that:

$$\sup_{F^{(\mathbb{N})}\in\tilde{\mathcal{F}}_U}\big\{J^*(F^{(\mathbb{N})},T)-J^{\pi}(F^{(\mathbb{N})},T)\big\} \geq C(\tilde{T}-1).$$

Proof. We fix preferences F^1 and F^τ and consider the induced preferences $F^{(\mathbb{N})}\in\tilde{\mathcal{F}}_U(F^1,F^\tau)$. Also, assume that there exist an admissible policy $\pi\in\mathcal{P}$ and a constant $\beta>0$ such that:

$$\min_{1\leq j\leq\tilde{T}-1}\{\mathcal{K}(\mathbb{P}^{\pi}_{\ell_{j+1}},\mathbb{P}^{\pi}_{\ell_j})\} > \beta.$$

In this case, the policy is guaranteed to explore sufficiently within each time segment. Hence, the environment can choose to change the preferences at the very last time period, i.e., $\tau=\ell_{\tilde{T}}=T$.

Therefore, the following sequence of inequalities holds:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_{\ell_{\tilde{T}}}}\left[J^*(F^{(\mathbb{N})},T) - \mathcal{J}^\pi(F^{(\mathbb{N})},T)\right] &= \mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=1}^{\ell_{\tilde{T}}-1}\big(r(S^*(F^1),F^1) - r(\psi_t(\mathcal{H}_{t-1}),F^1)\big)\Big] \\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_{\ell_{\tilde{T}}}}\Big[\sum_{t=\ell_{\tilde{T}}}^{T}\big(r\left(S^*(F^\tau),F^\tau\right) - r\left(\psi_t(\mathcal{H}_{t-1}),F^\tau\right)\big)\Big] \\
&\overset{(a)}{\geq} \mathbb{E}_{\mathbb{P}^\pi_{\ell_{\tilde{T}}}}\Big[\sum_{t=\ell_1}^{\ell_{\tilde{T}}-1}\big(r(S^*(F^1),F^1) - r(\psi_t(\mathcal{H}_{t-1}),F^1)\big)\Big] \\
&= \sum_{i=1}^{\tilde{T}-1}\mathbb{E}_{\mathbb{P}^\pi_{\ell_{\tilde{T}}}}\Big[\sum_{t=\ell_i}^{\ell_{i+1}-1}\big(r\left(S^*(F^1),F^1\right) - r\left(\psi_t(\mathcal{H}_{t-1}),F^1\right)\big)\Big] \\
&\overset{(b)}{=} \sum_{i=1}^{\tilde{T}-1}\mathbb{E}_{\mathbb{P}^\pi_{\ell_{i+1}}}\Big[\sum_{t=\ell_i}^{\ell_{i+1}-1}\big(r\left(S^*(F^1),F^1\right) - r\left(\psi_t(\mathcal{H}_{t-1}),F^1\right)\big)\Big] \\
&\overset{(c)}{\geq} \gamma\sum_{i=1}^{\tilde{T}-1}\mathbb{E}_{\mathbb{P}^\pi_{\ell_{i+1}}}\Big[\sum_{t=\ell_i}^{\ell_{i+1}-1}\mathbf{1}\left[\|S^*(F^1) - \psi_t(\mathcal{H}_{t-1})\|_1 \geq 1\right]\Big] \\
&\overset{(d)}{\geq} \gamma\sum_{i=1}^{\tilde{T}-1}\frac{\mathcal{K}(\mathbb{P}^\pi_{\ell_{j+1}},\mathbb{P}^\pi_{\ell_j})}{\mathcal{K}\left(F^1\,,\,F^\tau\right)} \geq \frac{\gamma\beta(\tilde{T}-1)}{\mathcal{K}\left(F^1\,,\,F^\tau\right)},
\end{aligned}$$

where (a) follows from the optimality of $S^*(F^\tau)$. For any given $i \in [\tilde{T}]$, the distribution of the purchase decision of customer $t \in \{\ell_i, \ldots, \ell_{i+1}-1\}$ is independent of the time at which the change occurs, provided it takes place after (or at) ℓ_{i+1}, which justifies (b). The inequality (c) follows from the definition of γ. Finally, (d) follows from Lemma 9; see (E.C.4).

Next, recall that by our initial assumption, the pre- and post-change preferences satisfy:

$$\sup\left\{\left|\log p_i(S,F^1) - \log p_i(S,F^\tau)\right| \,:\, \forall\, i \in S \cup \{0\},\, \forall\, S \in \mathcal{S}\right\} \leq \phi,$$

which indicates the maximum KL divergence is bounded above by ϕ. Accordingly, we have that:

$$\max\left\{\mathcal{K}(F^1\,,\,F^\tau\,;\,S) \,:\, S \in \mathcal{S}\right\} \leq \phi.$$

Finally, we obtain the following inequality on the difference between the expected revenue obtained by the oracle and by the policy π:

$$\sup_{F^{(\mathbb{N})}\in\tilde{\mathcal{F}}_U}\mathbb{E}_{\mathbb{P}^\pi_{\ell_{\tilde{T}}}}\left[J^*(F^{(\mathbb{N})},T) - \mathcal{J}^\pi(F^{(\mathbb{N})},T)\right] \geq C(\tilde{T}-1),$$

where $C \equiv C(\gamma,\phi,\beta) := \frac{\gamma\beta}{\phi}$. Therefore, we conclude the proof. ■

We now present the proof of Proposition 5. The proof follows from the observation that for any given $\beta > 0$, the conditions of either Lemma 1 or Lemma 2 must be satisfied. The proposition's conclusion therefore follows directly from the application of the relevant lemma.

Proof of Proposition 5. For $T \geq 2$, we define $\Delta = \lceil T^{1/2} \rceil$. Consequently, we have $\Delta \geq T^{1/2}$ and $\tilde{T} - 1 \geq \frac{1}{6}\sqrt{T}$. Next, we fix $\beta = 1$. For any policy, we apply either Lemma 1 or Lemma 2, depending on whether the policy explores sufficiently or not.

As a result, the following inequality holds:

$$\sup_{F^{(\mathbb{N})} \in \tilde{\mathcal{F}}_U} \left\{ J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) \right\} \geq \min \left\{ C(\gamma, 1)\Delta, C(\gamma, \phi, 1)(\tilde{T} - 1) \right\} \geq C(\gamma, \phi)\sqrt{T},$$

where $C(\gamma, \phi) = \min \left\{ C(\gamma, 1), \frac{1}{6} C(\gamma, \phi, 1) \right\}$. Therefore, we conclude the proof. ■

We now present the proof for the upper bound on the regret achieved by the active monitoring strategy described in Algorithm 4. Formally, we provide the proof of Proposition 6.

Proof of Proposition 6. Let $T \geq 2$, a pair (F^1, F^τ) be such that the induced preferences $F^{(\mathbb{N})}$ belong to $\mathcal{F}_U(F^1, F^\tau)$ so that they cannot be distinguished at $S^*(F^1)$. Define $\alpha \equiv (\alpha_{\mathrm{I}}, \alpha_{\mathrm{II}})$ as the two levels of control for the probability of the Type I and II errors, respectively. Furthermore, we define $D \equiv D(\alpha)$ as the smallest constant satisfying the following inequalities:

$$D(\alpha) \geq \max \left\{ 1, -\log(\alpha_{\mathrm{I}})(2\log(2))^{-1} \right\} \cdot \mathcal{K}\big(F^1 \,,\, F^\tau \,;\, S^*(F^1)\big)^{-2},$$
$$D(\alpha) \geq \max \left\{ 1/2, -\log(\alpha_{\mathrm{II}}/2)(2\log(2))^{-1} \right\} \cdot \mathcal{K}\big(F^\tau \,,\, F^1 \,;\, S^*(F^1)\big)^{-2}.$$

Next, we fix some $S \in \mathcal{S}$ such that $\mathcal{K}(F^1 \,,\, F^\tau \,;\, S) > 0$, which is used as an input for π, the active monitoring strategy, which is described in Algorithm 4. We denote by $\hat{k}$ the stopping rule that is used within the policy to detect the change. Specifically:

$$\hat{k}_{\ell_j}(\mathcal{H}_{\ell_j - 1}) := \ell_{j+1} \mathbf{1}(\hat{\Lambda}_{\ell_j} < 0),$$

where $\ell_0 = 1$, and $\ell_{j+1} = \ell_j + \Delta_o + \Delta_e$ (recall that $\Delta_o = D\sqrt{T}$ and $\Delta_e = D\log T$), for $j \in [\tilde{T} - 1]$. Also, we define $\tilde{T} := \lceil T/\Delta \rceil$, for $\Delta = \Delta_o + \Delta_e$. Then, we denote by j^* the index such that $\ell_{j^*} < \tau \leq \ell_{j^*+1}$, where $\ell_{\tilde{T}+1} = +\infty$ by convention. We divide the proof into 7 smaller steps.

Step 1. In the following, we denote the assortment strategy obtained through the policy by ψ_t, for $t \in [T]$. We omit the dependence of the policy on the filtration $(\mathcal{H}_t)_{t=1}^T$. Then, we derive the following sequence of inequalities for the regret of the policy:

$$J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) = \mathbb{E}_{\mathbb{P}^\pi_\tau} \Big[\sum_{t=1}^{\tau-1} \big(r(S^*(F^1), F^1) - r(\psi_t, F^1) \Big]$$
$$+ \mathbb{E}_{\mathbb{P}^\pi_\tau} \Big[\sum_{t=\tau}^{T} \big(r(S^*(F^\tau), F^\tau) - r(\psi_t, F^\tau) \big) \Big]$$

$$
\begin{aligned}
&= \mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=1}^{\tau-1}\big(r\left(S^*(F^1),F^1\right)-r\left(\psi_t,F^1\right)\big)\mathbf{1}\big(\psi_t\in\{S,S^*(F^\tau)\}\big)\Big]\\
&\quad+\mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=\tau}^{T}\big(r\big(S^*(F^\tau),F^\tau\big)-r\big(\psi_t,F^\tau\big)\big)\mathbf{1}\left(\psi_t\in\{S,S^*(F^1)\}\right)\Big]\\
&\le \delta\cdot\mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=1}^{T}\mathbf{1}\big(\psi_t=S\big)\Big]+\delta\cdot\mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=1}^{\tau-1}\mathbf{1}\big(\psi=S^*(F^\tau)\big)\Big]\\
&\quad+\delta\cdot\mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=\tau}^{T}\mathbf{1}\big(\psi=S^*(F^1)\big)\Big],\\
&\le \delta\cdot\mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=1}^{T}\mathbf{1}\big(\psi_t=S\big)\Big]+\delta\cdot\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k}-\tau)^+\big]+\delta\cdot\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big].
\end{aligned}
$$

Step 2. We begin by providing an upper bound to $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big]$. Next, for $j\in[j^*]$, we denote by $q_{f,j}$ the probability of a false alarm (i.e., the Type I error) at time $t=\ell_j+1$. Then, the following sequence of inequalities holds:

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big] &= \sum_{u=1}^{\tau-1}\mathbb{P}^\pi_\tau\big[(\tau-\hat{k})^+\ge u\big]\\
&= \sum_{u=1}^{\tau-1}\mathbb{P}^\pi_\tau\big[\hat{k}\le\tau-u\big]\\
&\overset{(a)}{\le} \sum_{j=1}^{j^*}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\mathbb{P}^\pi_\tau\Big[\bigcup_{m=1}^{j}\{\hat{k}=\ell_m+1\}\Big]\\
&\le \sum_{j=1}^{j^*}\sum_{i=\ell_j}^{\ell_{j+1}-1}\sum_{m=1}^{j}\mathbb{P}^\pi_\tau\big[\hat{k}=\ell_m+1\big]\overset{(b)}{\le}\sum_{j=1}^{j^*-1}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\sum_{m=1}^{j}q_{f,j}=\sum_{j=1}^{j^*-1}\left(\Delta_o+\Delta_e\right)jq_{f,j},
\end{aligned}
$$

where (a) follows from that $\tau\le\ell_{j^*+1}$ and the definition of our stopping-time random variable. Then, (b) follows from the definition of the Type I error and the fact that $\ell_{j^*}<\tau\le\ell_{j^*+1}$.

Step 3. Next, we derive an upper bound for the probability of the Type I error $q_{f,j}$, for $j\in[\tilde{T}-1]$. To proceed, we first fix some index $j\in[\tilde{T}-1]$. Then, assume that the following purchase decisions $Z^{\ell_{j+1}-\Delta_e-1},\ldots,Z^{\ell_{j+1}-1}$ for customers $\ell_{j+1}-\Delta_e-1$ to $\ell_{j+1}-1$ are available. Given these purchase decisions, we consider the following two statistical hypothesis:

$$
\begin{aligned}
H_{0,j}: Z^{\ell_{j+1}-\Delta_e-1},\ldots,Z^{\ell_{j+1}-1} &\sim F^1\left(\cdot\mid S\right),\\
H_{1,j}: Z^{\ell_{j+1}-\Delta_e-1},\ldots,Z^{\ell_{j+1}-1} &\sim F^\tau\left(\cdot\mid S\right).
\end{aligned}
$$

Next, we define the normalized log-likelihood ratio test $\hat{\Lambda}_{\ell_j}$ as follows:

$$
\hat{\Lambda}_{\ell_j}:=\frac{1}{\Delta_e}\sum_{u=\ell_{j+1}-\Delta_e-1}^{\ell_{j+1}-1}\log\left(\frac{F^1(Z^u\mid S)}{F^\tau(Z^u\mid S)}\right),
$$

which is guaranteed to be well-defined by the definition of $\mathcal{F}$.

Then, by the definition of probability of the Type I error, we have that $q_{f,j} := \mathbb{P}[\hat{\Lambda}_{\ell_j} < 0 \mid H_{0,j}]$. Moreover, if we condition on the event that $H_{0,j}$ is true, then we obtain the following equation for the expected value of the log-likelihood test:

$$\mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] = \mathbb{E}_{F^1}\Big[\frac{1}{\Delta_e}\sum_{u=\ell_{j+1}-\Delta_e-1}^{\ell_{j+1}-1} \log\left(\frac{F^1\left(Z^u \mid S\right)}{F^\tau\left(Z^u \mid S\right)}\right) \mid S\Big] = \mathcal{K}\left(F^1\,,\, F^\tau\,;\, S\right).$$

Consequently, we obtain the following sequence of inequalities:

$$\begin{aligned}
q_{f,j} &= \mathbb{P}\big[\hat{\Lambda}_{\ell_j} - \mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] < -\mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] \mid H_{0,\ell}\big] \\
&\leq \mathbb{P}\Big[\frac{1}{\Delta_e}\sum_{u=\ell_{j+1}-\Delta_e-1}^{\ell_{j+1}-1} \log\left(\frac{F^1\left(Z^u \mid S\right)}{F^\tau\left(Z^u \mid S\right)}\right) - \mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] \leq -\mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] \mid H_{0,j}\Big] \\
&\overset{(a)}{\leq} \exp(-2\Delta_e\big(\mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}]\big)^2) = \exp(-2\Delta_e\mathcal{K}(F^1\,,\, F^\tau\,;\, S)^2),
\end{aligned}$$

where (a) follows from the Hoeffding's inequality.

Therefore, are able to derive the following chain of inequalities:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau - \hat{k})^+\big] &\leq (\Delta_e + \Delta_o)\,\frac{j^*\,(j^*-1)}{2}\exp\big(-2\Delta_e\mathcal{K}(F^1\,,\, F^\tau\,;\, S)^2\big) \\
&\leq (\Delta_e + \Delta_o)\,\frac{\tilde{T}(\tilde{T}-1)}{2}\exp\big(-2\Delta_e\mathcal{K}(F^1\,,\, F^\tau\,;\, S)^2\big) \\
&\leq \frac{T^2}{2(\Delta_e+\Delta_o)}\exp\big(-2\Delta_e\mathcal{K}(F^1\,,\, F^\tau\,;\, S)^2\big) \\
&\leq \frac{T^2}{2D(\log(T)+\sqrt{T})}T^{-2C\mathcal{K}(F^1\,,\, F^\tau\,;\, S)^2} \overset{(a)}{\leq} \frac{T}{2D(\log(T)+\sqrt{T})} \leq \frac{1}{4D}\sqrt{T},
\end{aligned}$$

where (a) follows from the definition of constant $D \equiv D(\alpha)$ as used in Algorithm 4.

Step 4. Next, we derive an upper bound for $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k}-\tau)^+\big]$. To proceed, we denote by the probability of the Type II error $q_{d,j}$ of the statistical test for the sub-segment j. Specifically, we obtain the following sequence of inequalities:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k}-\tau)^+\big] = \sum_{u=0}^{T-\tau+1}\mathbb{P}^\pi_\tau\big[\hat{k} \geq \tau + u\big] &\leq \sum_{j=j^*}^{\tilde{T}-1}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\mathbb{P}^\pi_\tau\Big[\bigcap_{m=j^*}^{j}\{\hat{k} \neq \ell_m + 1\}\Big] \\
&\overset{(a)}{\leq} \sum_{j=j^*+2}^{\tilde{T}-1}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\mathbb{P}^\pi_\tau\Big[\bigcap_{m=j^*+2}^{j}\{\hat{k} \neq i_m + 1\}\Big] + 3\,(\Delta_e + \Delta_o) \\
&\overset{(b)}{=} (\Delta_e + \Delta_o)\,\Big(3 + \sum_{j=j^*+2}^{\tilde{T}-1}\left(q_{d,j^*+2}\right)^{j-j^*-1}\Big) \\
&= (\Delta_e + \Delta_o)\,\Big(3 + q_{d,j^*+2}\frac{1-\left(q_{d,j^*+2}\right)^{\tilde{T}-j^*}}{1-q_{d,j^*+2}}\Big) \leq \frac{3\,(\Delta_e + \Delta_o)}{1-q_{d,j^*+2}},
\end{aligned}$$

where (a) follows from the change could be anywhere within segment $\{\ell_{j^*}, \ldots, \ell_{j^*+1}-1\}$. Moreover, (b) holds since $q_{d,j} = q_{d,j^*+2}$, for all $j \in \{j^*+2, \ldots, \tilde{T}\}$. Indeed, the probability of the Type II error only depends on the fact that the change happens, but not when it happens.

Step 5. Next, we derive an upper bound for the probability of the Type II error. To proceed, we fix some index $j \in \{j^*+2, \ldots, \tilde{T}-1\}$. Then, by the definition of the Type II error, we have that $q_{d,j} = \mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} \geq 0\big]$. Moreover, we use similar arguments as earlier, and, in particular, the Hoeffding's inequality, to obtain the following upper bound:

$$\mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} \geq 0\big] = \mathbb{P}\big[\hat{\Lambda}_{\ell_j} - \mathbb{E}_{H_{1,j}}[\hat{\Lambda}_{\ell_j}] \geq -\mathbb{E}_{H_{1,j}}[\hat{\Lambda}_{\ell_j}] \,|\, H_{1,j}\big] \leq 2\exp(-2\Delta_e \mathcal{K}(F^\tau\,,\, F^1\,;\, S)^2).$$

Therefore, we arrive at the following inequality:

$$\frac{1}{1-q_{d,j}} \leq \big[1 - 2\exp(-2\Delta_e \mathcal{K}(F^\tau\,,\, F^1\,;\, S)^2)\big]^{-1}.$$

Consequently, we can derive an upper bound for the detection delay. Formally, we obtain:

$$\mathbb{E}_{\mathbb{P}^\pi_\tau}[(\hat{k}-\tau)^+] \leq 3(\Delta_e + \Delta_o)\big[1 - 2\exp(-2\Delta_e \mathcal{K}(F^\tau\,,\, F^1\,;\, S)^2)\big]^{-1} \overset{(a)}{\leq} 3\,(\Delta_e + \Delta_o) \leq 6D\sqrt{T},$$

where (a) follows from the choice of constant D.

Step 6. We now establish the final upper bound necessary for the proof. Specifically, we have:

$$\sum_{t=1}^{T} \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\mathbf{1}(\psi_t = S)\big] \leq (\tilde{T}-1)\Delta_e \leq \frac{\Delta_e T}{\Delta_e + \Delta_o} \leq \frac{T\log T}{\sqrt{T} + \log T} \leq \sqrt{T}\log T,$$

which provides the desired bound on the summation term.

Step 7. With all required bounds in place, we now bound the difference between the expected revenue achieved by the oracle and the expected revenue of our policy. Specifically, we have:

$$\begin{aligned} J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) &\leq \delta\big(\sqrt{T}\log T + \frac{1}{4D}\sqrt{T} + 6D\sqrt{T}\big) \\ &= \delta\sqrt{T}\log T\big(1 + \frac{1}{4D\log(T)} + 6D\frac{1}{\log(T)}\big) \leq \delta\sqrt{T}\log T\big(1 + \frac{1}{4D} + 6D\big), \end{aligned}$$

where we obtain the desired result by defining some constant $C \equiv C(\delta, \alpha, S) > 0$. Specifically, if we define $C := \delta(1 + (4D(\alpha))^{-1} + 6D(\alpha)\big)$, then we conclude the proof. ■

E.C.3.2 Proofs for Appendix A.3.2

Next, we establish a lower bound on achievable performance and an upper bound on the regret attained by Algorithm 5. Specifically, we present proofs for Proposition 7 and 8, which rely on two technical results, namely, Lemmas 3 and 4 that are proved in (E.C.3.3).

Proof of Proposition 7. We fix F^1 and F^τ such that the induced preferences $F^{(\mathbb{N})}$ belong to $\mathcal{F}_A$. Let $\pi \in \mathcal{P}$ represent a non-anticipatory policy characterized by the assortment mapping $\psi_t(\mathcal{H}_{t-1}) \in \mathcal{S}$

for each $t \in [T]$. The random vector of consumer purchasing decisions is defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$. Given the filtration $(\mathcal{H}_t)_{t=0}^T$, the random variable $\mathcal{J}^\pi(F^{(\mathbb{N})}, T)$ is similarly defined on this space (refer to the discussion in Lemma 5). For an arbitrary $\eta > 0$, we define:

$$B_\eta := \Big\{\omega \in \Omega \ : \ J^*(F^{(\mathbb{N})}, T) - \mathcal{J}^\pi(F^{(\mathbb{N})}, T) < \eta. \Big\}.$$

We define $j_0 := \lceil \eta/\gamma \rceil$. Moreover, we introduce $\hat{k}_1, \dots \hat{k}_T$, which are defined as follows:

$$\hat{k}_1 := \min\big\{1 \le t \le T \ : \ \{\psi_t(\mathcal{H}_{t-1}) = S^*(F^\tau)\} \cup \{t = T\}\big\},$$

and, for $i \ge 1$:

$$\hat{k}_{i+1} := \begin{cases} \min\big\{\hat{k}_i < t \le T \ : \ \{\psi_t(\mathcal{H}_{t-1}) = S^*(F^\tau)\} \cup \{t = T\}\big\}, & \text{if } \hat{k}_i < T, \\ T, \text{ if } \hat{k}_i \ge T. \end{cases}$$

Next, we define the following stopping rule $\hat{k}^* := \hat{k}_{j_0}$ to estimate the change-time τ.

Lemma 3. *For any $\omega \in B_\eta$, we have: $0 \le \hat{k}^* - \tau \le 2j_0$.*

By leveraging Lemma 3, we establish the following chain of set inclusions:

$$B_\eta \subseteq \big\{\omega \in \Omega \ : \ 0 \le \hat{k}^* - \tau \le 2j_0\big\} \subseteq \big\{\omega \in \Omega \ : \ \big|\hat{k}^* - \tau\big| \le 2j_0\big\}.$$

Consequently, the following inequality is valid:

$$\mathbb{P}_\tau^\pi\big[B_\eta\big] \le \mathbb{P}_\tau^\pi\big[\big|\hat{k}^* - \tau\big| \le 2j_0\big],$$

and, by considering the complementary of B_η, denoted by B_η^c, we obtain the following inequality:

$$\mathbb{P}_\tau^\pi\big[B_\eta^c\big] \ge \mathbb{P}_\tau^\pi\big[\big|\hat{k}^* - \tau\big| > 2j_0\big] \quad \forall \tau \in [T+1].$$

Hence, since the former inequality holds for any $\tau \in [T+1]$, the following inequality holds:

$$\sup_{1 \le \tau \le T+1} \mathbb{P}_\tau^\pi\big[B_\eta^c\big] \ge \sup_{1 \le \tau \le T+1} \mathbb{P}_\tau^\pi\big[\big|\hat{k}^* - \tau\big| > 2j_0\big].$$

Lemma 4. *There exists $\tilde{C} \equiv \tilde{C}(\vartheta) > 0$ and $\alpha(\vartheta) > 0$, then, any admissible stopping rule $\hat{k}$ with respect to the history $(\mathcal{H}_{t-1})_{t=1}^T$ must satisfy:*

$$\sup_{1 \le \tau \le T} \mathbb{P}_\tau^\pi\big[\big|\hat{k} - \tau\big| > \tilde{C} \log T\big] \ge \alpha.$$

We fix $\eta := \big(C_1 \log T - \gamma\big)^+$, where $C_1 = \frac{\tilde{C}\gamma}{2}$, and $\tilde{C} \equiv \tilde{C}(\vartheta)$ is the constant from Lemma 4. Then, we derive the following sequence of inequalities:

$$2j_0 = 2\lceil \eta/\gamma \rceil = 2\lceil \frac{\big(C_1 \log T - \gamma\big)^+}{\gamma} \rceil = 2\lceil \big(\frac{\tilde{C}}{2} \log T - 1\big)^+ \rceil \le \tilde{C} \log T.$$

We can now establish a sequence of inequalities that lead to our main result. First, applying our previous findings and Lemma 4, we obtain:

$$\sup_{1\le\tau\le T}\mathbb{P}^\pi_\tau\big[B^c_\eta\big]\ge\sup_{1\le\tau\le T}\mathbb{P}^\pi_\tau\big[\big|\hat{k}-\tau\big|>2j_0\big]\ge\sup_{1\le\tau\le T}\mathbb{P}^\pi_\tau\big[\big|\hat{k}-\tau\big|>\tilde{C}\log T\big]\ge\alpha.$$

This chain of inequalities demonstrates that the probability of the complement of B_η is bounded below by α, which plays an important role for establishing our regret bound. Formally:

$$\sup_{F^{(\mathbb{N})}\in\mathcal{F}(F^1,F^\tau)}\big\{J^*(F^{(\mathbb{N})},T)-J^\pi(F^{(\mathbb{N})},T)\big\}\ge\sup_{1\le\tau\le T}\eta\mathbb{P}^\pi_\tau\big[B^c_\eta\big]\ge\sup_{1\le\tau\le T}\alpha\big(C_1\log T-\gamma\big)^+\overset{(a)}{=}C\log T,$$

where equality (a) holds for all $T\ge\exp\left(\frac{4}{C}\right)$ for $C\equiv C(\gamma,\vartheta):=\alpha\gamma\frac{\tilde{C}}{4}$.

Both constants C and α depend only on parameters γ and ϑ, and are independent of the specific choice of preferences F^1 and F^τ. This observation completes the proof. ■

Next, we turn our attention to deriving an upper bound on the regret achieved by the passive-monitoring strategy as formally defined in Algorithm 5.

Proof of Proposition 8. Assume that $T\ge 2$, F^1 and F^τ are such that the preferences $F^{(\mathbb{N})}$ they induce belong to $\mathcal{F}_D$. Define $\alpha\equiv(\alpha_{\rm I},\alpha_{\rm II})$ as the two levels of control for the Type I and II errors, respectively. We define $D\equiv D(\alpha)$ as the smallest constant satisfying the following inequalities:

$$D(\alpha)\ge\max\big\{1,-\log(\alpha_{\rm I})(2\log(2))^{-1}\big\}\mathcal{K}\big(F^1\,,\,F^\tau\,;\,S^*(F^1)\big)^{-2},$$
$$D(\alpha)\ge\max\big\{1,-\log(\alpha_{\rm II}/2)(2\log(2))^{-1}\big\}\mathcal{K}\big(F^\tau\,,\,F^1\,;\,S^*(F^1)\big)^{-2}.$$

We define the customer batch size, denoted by $\Delta:=D\log T$, as specified in Algorithm 5. Let $\ell_1:=1$, and define $\ell_{j+1}:=\ell_j+\Delta$ for $j\in[\tilde{T}-1]$, with $\ell_{\tilde{T}}:=T$, where $\tilde{T}:=\lceil T/\Delta\rceil$. Additionally, we denote by j^* the index such that $\ell_{j^*}<\tau\le\ell_{j^*+1}$, where $\ell_{\tilde{T}+1}:=\infty$ by convention. For the statistical test used in policy π, we introduce $\hat{\Lambda}_{\ell_j}$ as the statistic (log-likelihood) computed for customers belonging to the time segment j. We also define $\hat{k}$ as the stopping rule employed in policy $\pi\equiv\pi(D,F^1,F^\tau)$ from Algorithm 5. Formally, the stopping rule is given by:

$$\hat{k}_{\ell_j}(\mathcal{H}_{\ell_j-1}):=\ell_{j+1}\mathbf{1}\big(\hat{\Lambda}_{\ell_j}<0\big).$$

Step 1. In the following, we denote by ψ_t the assortment strategy for $t\in[T]$ corresponding to the policy π. For clarity, we omit the dependence of the policy on the filtration $(\mathcal{H}_t)_{t=0}^T$. We then bound the difference in the expected revenue between the oracle and our policy in terms of the proposed stopping rule. Specifically, we derive the following sequence of inequalities:

$$J^*(F^{(\mathbb{N})},T)-J^\pi(F^{(\mathbb{N})},T)=\mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=1}^{\tau-1}\big(r(S^*(F^1),F^1)+r(\psi_t,F^1)\big)\Big]$$

$$
\begin{aligned}
&- \mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=\tau}^{T}\big(r(S^*(F^\tau),F^\tau)+r(\psi_t,F^\tau)\big)\Big]\\
&= \mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=1}^{\tau-1}\big(r(S^*(F^1),F^1)-r\big(S^*(F^\tau),F^1\big)\big)\mathbf{1}\big(\psi_t=S^*(F^\tau)\big)\Big]\\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\Big[\sum_{t=\tau}^{T}\big(r(S^*(F^\tau),F^\tau)-r(S^*(F^1),F^\tau)\big)\mathbf{1}\big(\psi_t=S^*(F^1)\big)\Big]\\
&\le \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\big(\hat{k}-\tau\big)^+\big]\big(r(S^*(F^\tau),F^\tau)-r(S^*(F^1),F^\tau)\big)\\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\big(\tau-\hat{k}\big)^+\big]\big(r(S^*(F^1),F^1)-r(S^*(F^\tau),F^1)\big).
\end{aligned}
$$

Consequently, we obtain the following upper bound for the difference between the expected revenue obtained by the oracle and the one obtained by policy π:

$$
J^*(F^{(\mathbb{N})},T)-J^\pi(F^{(\mathbb{N})},T)\le \delta\big(\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k}-\tau)^+\big]+\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big]\big).
$$

In the following steps, we derive an upper bound for both $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\hat{k}-\tau)^+\big]$ and $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big]$. Then, we derive the appropriate upper bound for the regret of policy π.

Step 2. We proceed to analyze the expected detection advance $\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\big(\tau-\hat{k}\big)^+\big]$ by relating it to the probability of the Type I error. Let $q_{f,j}$ denote the probability of a false alarm at time $t=\ell_j+1$ for each $j\in[j^*]$. We establish the following sequence of inequalities:

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau-\hat{k})^+\big] &= \sum_{u=1}^{\tau-1}\mathbb{P}^\pi_\tau\big[(\tau-\hat{k})^+\ge u\big]=\sum_{u=1}^{\tau-1}\mathbb{P}^\pi_\tau\big[\hat{k}\le\tau-u\big]\\
&\overset{(a)}{\le} \sum_{j=1}^{j^*}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\mathbb{P}^\pi_\tau\Big[\bigcup_{m=1}^{j}\big\{\hat{k}=\ell_m+1\big\}\Big]\\
&\le \sum_{j=1}^{j^*}\sum_{i=\ell_j}^{\ell_{j+1}-1}\sum_{m=1}^{j}\mathbb{P}^\pi_\tau\big[\hat{k}=\ell_m+1\big]\overset{(b)}{\le}\sum_{j=1}^{j^*-1}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\sum_{m=1}^{j}q_{f,j}=\sum_{j=1}^{j^*-1}\Delta j q_{f,j},
\end{aligned}
$$

where (a) follows from that $\tau\le\ell_{j^*+1}$ and the definition of our stopping-time random variable. Then, (b) follows from the definition of the Type I error and the fact that $\ell_{j^*}<\tau\le\ell_{j^*+1}$.

Step 3. We find a bound for $q_{f,j}$, for all $j\in[\tilde{T}-1]$. To begin, we fix $j\in[\tilde{T}-1]$ and assume that the purchase decisions $Z^{\ell_j},\ldots,Z^{\ell_{j+1}-1}$ are available. We consider the hypothesis test:

$$
\begin{aligned}
H_{0,j}&: Z^{\ell_j},\ldots,Z^{\ell_{j+1}-1}\ \sim\ F^1\,|\,S^*(F^1),\\
H_{1,j}&: Z^{\ell_j},\ldots,Z^{\ell_{j+1}-1}\ \sim\ F^\tau\,|\,S^*(F^1).
\end{aligned}
$$

Moreover, we define the normalized log-likelihood ratio test $\hat{\Lambda}_{\ell_j}$ as follows:

$$
\hat{\Lambda}_{\ell_j}:=\frac{1}{\Delta}\sum_{u=\ell_j}^{\ell_{j+1}-1}\big(\log F^1\big(Z^u\mid S^*(F^1)\big)-\log F^\tau\big(Z^u\mid S^*(F^1)\big)\big),
$$

which is well-defined by definition of $\mathcal{F}$.

Moreover, by definition of the probability of the Type I error, we have that: $q_{f,j} := \mathbb{P}[\hat{\Lambda}_{\ell_j} < 0 \mid H_{0,j}]$. This expression corresponds to the probability of rejecting the null hypothesis when it is assumed to be true. Next, if we assume that $H_{0,j}$ is true, then we have:

$$\begin{aligned}\mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] &= \mathbb{E}_{F^1}\big[\frac{1}{\Delta}\sum_{u=\ell_j}^{\ell_{j+1}-1}\big(\log F^1\big(Z^u \mid S^*(F^1)\big) - \log F^\tau\big(Z^u \mid S^*(F^1)\big)\big) \mid S^*(F^1)\big]\\ &= \mathcal{K}(F^1\,,\,F^\tau\,;\,S^*(F^1)).\end{aligned}$$

Therefore, we obtain the following sequence of inequalities:

$$\begin{aligned} q_{f,j} &= \mathbb{P}\big[\hat{\Lambda}_{\ell_j} - \mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] < -\mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}] \mid H_{0,\ell}\big]\\ &\overset{(a)}{\leq} \exp\big(-2\Delta\big(\mathbb{E}_{H_{0,j}}[\hat{\Lambda}_{\ell_j}]\big)^2\big)\\ &= \exp\big(-2\Delta\mathcal{K}(F^1\,,\,F^\tau;S^*(F^1))^2\big) \overset{(b)}{\leq} \exp\big((\log(\alpha_{\mathrm{I}})/\log(2))\log T\big) = T^{-2}\alpha_{\mathrm{I}}^{-\frac{1}{2\log 2}} \overset{(c)}{\leq} \alpha_{\mathrm{I}}^{\frac{\log T}{\log 2}} \leq \alpha_{\mathrm{I}},\end{aligned}$$

where (a) follows from the Hoeffding's inequality. Then, (b) follows by the definition of $\Delta \equiv D(\alpha)\log T$, and (c) follows from that $\log T/\log 2 \geq 1$ for all $T \geq 2$. Also, as a side observation, our test is guaranteed to control the Type I error at level α_{I} for any $T \geq 2$.

Therefore, we conclude that:

$$\mathbb{E}_{\mathbb{E}_\tau^\pi}\big[(\tau - \hat{k})^+\big] \leq \Delta\frac{j^*(j^*-1)}{2}T^{-2}\alpha_{\mathrm{I}}^{-\frac{1}{2\log 2}} \leq \Delta\frac{\tilde{T}(\tilde{T}-1)}{2}T^{-2}\alpha_{\mathrm{I}}^{-\frac{1}{2\log 2}} \leq \frac{1}{2\Delta}\alpha_{\mathrm{I}}^{-\frac{1}{2\log 2}}.$$

Step 4. We proceed to analyze the expected detection delay $\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k}-\tau)^+\big]$ by relating it to the probability of the Type II errors in our sequential testing procedure. Given the probability of the Type II error $q_{d,j}$ corresponding to the statistical test from policy π at within sub-segment j, we obtain the following upper bound:

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[\big(\hat{k}-\tau\big)^+\big] &= \sum_{u=0}^{T-\tau+1}\mathbb{P}_\tau^\pi\big[\hat{k} \geq \tau + u\big]\\ &\leq \sum_{j=j^*}^{\tilde{T}-1}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\mathbb{P}_\tau^\pi\big[\bigcap_{m=j^*}^{j}\big\{\hat{k} \neq \ell_m + 1\big\}\big]\\ &\overset{(a)}{\leq} \sum_{j=j^*+2}^{\tilde{T}-1}\sum_{\ell=\ell_j}^{\ell_{j+1}-1}\mathbb{P}_\tau^\pi\big[\bigcap_{m=j^*+2}^{j}\big\{\hat{k} \neq \ell_m + 1\big\}\big] + 3\Delta\\ &\overset{(b)}{=} \Delta\big(3 + \sum_{j=j^*+2}^{\tilde{T}-1}\big(q_{d,j^*+2}\big)^{j-j^*-1}\big) = \Delta\big(3 + q_{d,j^*+2}\frac{1-\big(q_{d,j^*+2}\big)^{\tilde{T}-j^*}}{1-q_{d,j^*+2}}\big) \leq \frac{3\Delta}{1-q_{d,j^*+2}},\end{aligned}$$

where (a) follows from the change could be anywhere within segment $\{\ell_{j^*},\ldots,\ell_{j^*+1}-1\}$. Moreover, (b) holds since $q_{d,j} = q_{d,j^*+2}$, for all $j \in \{j^*+2, \ldots, \tilde{T}\}$. Indeed, the probability of the Type II error only depends on the fact that the change happens.

Step 5. This step consists in finding an upper bound for the probability of the Type II error. We fix some index $j \in \{j^* + 2, \ldots, \tilde{T} - 1\}$. Then, the probability of the Type II error is given by:

$$q_{d,j} := \mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} \geq 0\big] = \mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} > 0\big] + \mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} = 0\big].$$

And, similarly as before, we use the Hoeffding's inequality and obtain the following upper bound for the first part of the Type II error probability:

$$\mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} > 0\big] = \mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} - \mathbb{E}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j}\big] > -\mathbb{E}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j}\big]\big] \leq \exp\big(-2\Delta\mathcal{K}(F^\tau\,,\,F^1\,;\,S^*(F^1))^2\big).$$

Using a similar approach to find an upper bound for $\mathbb{P}_{H_{1,j}}\big[\hat{\Lambda}_{\ell_j} = 0\big]$, we obtain that:

$$q_{d,j} \leq 2\exp\big(-2\Delta\mathcal{K}(F^\tau\,,\,F^1\,;\,S^*(F^1))^2\big) \leq \alpha_{\mathrm{II}}.$$

Therefore, the following inequality is satisfied:

$$\frac{1}{1 - q_{d,j}} \leq (1 - \alpha_{\mathrm{II}})^{-1}.$$

Consequently, we obtain the following upper bound for the error made by the stopping time $\hat{k}$:

$$\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k} - \tau)^+\big] \leq 3\Delta(1 - \alpha_{\mathrm{II}})^{-1}.$$

Finally, we conclude that the difference in the expected revenue between the oracle strategy and π is bounded above as follows:

$$\begin{aligned} J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) &\leq \delta\big(\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k} - \tau)^+\big] + \mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\tau - \hat{k})^+\big]\big) \\ &\leq \delta\big(\frac{1}{2\Delta}\alpha_{\mathrm{I}}^{-\frac{1}{2\log 2}} + \frac{3\Delta}{1 - \alpha_{\mathrm{II}}}\big) \leq C_1 + C_2 \log T, \end{aligned}$$

where $C_1 := C_1(\delta, \alpha_{\mathrm{I}}) = \delta(2\log 2)^{-1}\alpha_{\mathrm{I}}^{-\frac{1}{2\log 2}}$ and $C_2 := C_2(\delta, \alpha_{\mathrm{II}}) = 3\delta(1 - \alpha_{\mathrm{II}})^{-1}$. Finally, as this constant is independent of F^1 and F^τ, by taking the supremum over all possible preferences $F^{(\mathbb{N})} \in \mathcal{F}(F^1, F^\tau)$, we obtain the desired upper bound for the minimax regret of the policy π. This observation, in turn, concludes the proof. ■

E.C.3.3 Proofs of Lemmas 3 and 4

In this section, we provide the proofs of some essential lemmas that are used within the proof of the above propositions. Specifically, we provide the proofs of Lemmas 3 and 4.

Proof of Lemma 3. We establish the statement using a proof by contradiction. Assume, for the sake of contradiction, that either $\hat{k}^* - \tau < 0$ or $\hat{k}^* - \tau > 2j_0$. We analyze these two cases separately.

Case 1: Suppose that $\hat{k}^* - \tau > 2j_0$. Then, the following inequality holds:

$$\sum_{t=\tau}^{\hat{k}^*} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) \neq S^*(F^\tau)\big) = \sum_{t=\tau}^{\hat{k}_{j_0}} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) \neq S^*(F^\tau)\big)$$

$$\overset{(a)}{\geq} \hat{k}_{j_0} - j_0 > \tau + j_0 \geq 1 + j_0 \geq j_0 \overset{(b)}{\geq} \frac{\eta}{\gamma},$$

where (a) follows from that $\hat{k}_{j_0} - \tau > 2j_0$, and that the policy offers at most j_0 times assortment $S^*(F^\tau)$ in the time horizon $[\hat{k}_{j_0}]$. Moreover, (b) follows from $j_0 = \lceil \eta/\gamma \rceil$.

Next, recall that $\omega \in B_\eta$. Consequently, the following sequence of inequalities hold:

$$\eta > J^*(F^{(\mathbb{N})}, T) - \mathcal{J}^\pi(F^{(\mathbb{N})}, T) \geq \gamma \sum_{t=\tau}^{T} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) \neq S^*(F^\tau)\big)$$

$$\geq \gamma \sum_{t=\tau}^{\hat{k}^*} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) \neq S^*(F^\tau)\big) > \gamma \frac{\eta}{\gamma} = \eta,$$

which is clearly a contradiction. Therefore, we must have $\hat{k}^* - \tau \leq 2j_0$.

Case 2. Assume that $\hat{k}^* - \tau < 0$. Then, the following sequence of inequalities holds:

$$\sum_{t=1}^{\hat{k}^*} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) \neq S^*(F^1)\big) \geq \sum_{t=1}^{\hat{k}^*} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) = S^*(F^\tau)\big) = \sum_{t=1}^{\hat{k}_{j_0}} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) = S^*(F^\tau)\big) \overset{(a)}{\geq} j_0 > \frac{\eta}{\gamma},$$

where (a) follows from the construction of the indices $\hat{k}_i$ for all $i \in [T]$.

Next, since $\omega \in B_\eta$, the following sequence of inequalities holds:

$$\eta > J^*(F^{(\mathbb{N})}, T) - \mathcal{J}^\pi(F^{(\mathbb{N})}, T) \geq \gamma \sum_{t=1}^{\tau-1} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) \neq S^*(F^1)\big)$$

$$\geq \gamma \sum_{t=1}^{\hat{k}^*} \mathbf{1}\big(\psi_t(\mathcal{H}_{t-1}) \neq S^*(F^1)\big) > \gamma \frac{\eta}{\gamma} = \eta,$$

which is clearly a contradiction. Therefore, we have that $0 \leq \hat{k}^* - \tau \leq 2j_0$, which, together with the first case considered, concludes the proof. ■

Proof of Lemma 4. Our proof closely follows the approach by Besbes and Zeevi (2011), which itself draws inspiration from Korostelev (1988). Let $\hat{k}$ denote any admissible stopping rule based on the filtration $(\mathcal{H}_t)_{t=1}^T$. Next, we define a measure of divergence between F^1 and F^τ as follows:

$$\phi\left(F^1, F^\tau\right) := \max\left\{\left|\log F^1(z|S) - \log F^\tau(z|S)\right| \;:\; z \in \{0,1\}^N,\ \|z\|_1 \leq 1,\ z_i = 0 \quad \forall\, i \notin S,\ S \in \mathcal{S}\right\},$$

which is well defined as $0 < \phi\left(F^1, F^\tau\right) < \vartheta$.

The lower bound $0 < \phi\left(F^1, F^\tau\right)$ follows from the definition of $\mathcal{F}_D$ (as preferences are passively detectable). Indeed, assume for the sake of contradiction that $\phi\left(F^1, F^\tau\right) = 0$. Then, we have that

$F^1(z \mid S) = F^\tau(z \mid S) = 1$ for all z and S as defined above. In particular, we have that:

$$\mathcal{K}\big(F^1\,,\,F^\tau\,;\,S^*(F^1)\big) = 0,$$

which contradicts the assumption that the preferences $F^{(\mathbb{N})}$ induced by F^1 and F^τ are in $\mathcal{F}_D$.

Next, we fix $\beta \in (0,1)$ arbitrarily, and define the constant $C = (2\vartheta)^{-1}$. Also, we introduce:

$$B(\vartheta, x) := e^{-\vartheta}\frac{x^{1-C\vartheta}}{2C\log(x)+2} = e^{-\vartheta}\frac{x^{\frac{1}{2}}}{2C\log(x)+2}.$$

which is increasing for x, and such that $\lim_{x\to+\infty} B(\vartheta, x) = +\infty$.

Next, given $\beta \in (0,1)$, we construct $n_0 > 0$ such that:

$$n_0 = \max\big\{j \in \mathbb{N}_{\geq 1}\ :\ B(\vartheta, j) < \beta^{-2}\big\}.$$

Then, we define $g(x) = x(C\log(x)+1)^{-1}$. Observe that $g(\cdot)$ is an increasing function and that $\lim_{x\to+\infty} g(x) = +\infty$. Moreover, we introduce:

$$n_1 = \max\big\{j \in \mathbb{N}_{\geq 1}\ :\ j \geq n_0,\ g(j) \geq 3/2\big\}.$$

Case 1. Assume that $T \geq n_1$. Define $\Delta = \lceil C\log T\rceil$ and $\tilde{T} = \lceil T/\Delta\rceil$. Then, observe that

$$T/\Delta = T\big(\lceil C\log T\rceil\big)^{-1} \geq T\big(C\log T + 1\big)^{-1} = g(T) \geq g(n_1) \geq 3/2,$$

which then implies that, $\tilde{T} \geq 2$.

Next, define $\ell_j = 1 + (j-1)\Delta$ for $j \in [\tilde{T}-1]$, and let $\ell_{\tilde{T}} = T$. Moreover, we denote by $Z := \big(Z^1, \ldots, Z^T\big)$ the random vector corresponding to the customer's purchase decisions over the T time period, which is defined over some probability space $(\Omega, \mathcal{B}, \mathbb{P})$. Next, we introduce a new random variable $\tilde{Z}_j$, for each $j \in [\tilde{T}]$, as follows:

$$\tilde{Z}_j = \sum_{t=\ell_j}^{\ell_{j+1}-1}\big(\log F^1\left(Z^t \mid \psi_t(\mathcal{H}_{t-1})\right) - \log F^\tau\left(Z^t \mid \psi_t(\mathcal{H}_{t-1})\right)\big).$$

The next step consist of showing that the following inequality holds:

$$\min_{1\leq j\leq \tilde{T}-1} \mathbb{P}^\pi_{\ell_j}\big[|\hat{k}-\tau| > \Delta/3\big] \geq 1-\beta.$$

Hence, we assume for the sake of contradiction that this inequality does not hold. That is:

$$\min_{1\leq j\leq \tilde{T}-1} \mathbb{P}^\pi_{\ell_j}\big[|\hat{k}-\tau| > \Delta/3\big] < 1-\beta.$$

Then, we define the event $A_j := \big\{\omega \in \Omega\ :\ |\hat{k}-\ell_j| \leq \Delta/3\big\}$ for $j \in [\tilde{T}-1]$. Note that the events A_j are disjoint as each segment $\{\ell_j, \ldots, \ell_{j+1}-1\}$ is of size Δ, and:

$$\big\{\omega \in \Omega\ :\ |\hat{k}-\ell_{\tilde{T}}| > \Delta/3\big\} \supset \bigcup_{j=1}^{\tilde{T}-1} A_j.$$

Thus, the following chain of inequalities hold:

$$\mathbb{P}^{\pi}_{\ell_{\tilde{T}}}\big[|\hat{k}-\tau|>\Delta/3\big] \geq \mathbb{P}^{\pi}_{\ell_{\tilde{T}}}\Big[\bigcup_{j=1}^{\tilde{T}-1} A_j\Big] = \sum_{j=1}^{\tilde{T}-1} \mathbb{P}^{\pi}_{\ell_{\tilde{T}}}\big[|\hat{k}-\ell_j|\leq \Delta/3\big] = \sum_{j=1}^{\tilde{T}-1} \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{\tilde{T}}}}\big[\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big].$$

Importantly, we have that $\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)$ is $\mathcal{H}_{\ell_{j+1}-1}$-measurable, and its distribution does not depend on time changes occurring after $\ell_{j+1}-1$. Hence, the following equality holds:

$$\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{\tilde{T}}}}\big[\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big] = \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{j+1}}}\big[\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big].$$

Next, for V, an $\mathcal{H}_{\ell_{j+1}-1}$-measurable random variable, we derive the following equalities:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_j}}\big[e^{\tilde{Z}_j}V\big] &= \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_j}}\Big[\prod_{t=\ell_j}^{\ell_{j+1}-1}\frac{F^1\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}\big)\big)}{F^\tau\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}\big)\big)}V\Big] = \int_\Omega \prod_{t=\ell_j}^{\ell_{j+1}-1}\frac{F^1\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)}{F^\tau\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)}V(\omega)\mathbb{P}^{\pi}_{\ell_j}(\omega)\,d\omega \\
&= \int_\Omega \prod_{t=\ell_j}^{\ell_{j+1}-1}\frac{F^1\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)}{F^\tau\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)}V(\omega) \\
&\quad\cdot\prod_{t=1}^{\ell_j-1}F^1\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)\prod_{t=\ell_j}^{T}F^\tau\big(Z^t\mid\psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)\,d\omega \\
&= \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{j+1}}}\big[V\big],
\end{aligned}$$

where we use a change of measure type of argument, together with Lemma 6.

Moreover, since $\tilde{Z}_j \geq -\Delta\vartheta$, the following sequence of inequalities holds:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_j}}\big[e^{\tilde{Z}_j}\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big] &\geq \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_j}}\big[e^{-\Delta\vartheta}\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big] \\
&\geq \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_j}}\big[e^{-\big(1+C\log T\big)\vartheta}\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big] = \frac{e^{-\vartheta}}{T^{C\vartheta}}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_j}}\big[\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big].
\end{aligned}$$

Consequently, we derive the following sequence of inequalities:

$$\begin{aligned}
\mathbb{P}^{\pi}_{\ell_{\tilde{T}}}\big[|\hat{k}-\tau|>\Delta/3\big] &\geq \sum_{j=1}^{\tilde{T}-1}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{\tilde{T}}}}\big[\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big] \\
&\geq \sum_{j=1}^{\tilde{T}-1}\frac{e^{-\vartheta}}{T^{C\vartheta}}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_j}}\big[\mathbf{1}\big(|\hat{k}-\ell_j|\leq \Delta/3\big)\big] \quad \geq \frac{(\tilde{T}-1)e^{-\vartheta}}{T^{C\vartheta}}\min_{1\leq j\leq \tilde{T}-1}\mathbb{P}^{\pi}_{\ell_j}\big[|\hat{k}-\ell_j|\leq \Delta/3\big].
\end{aligned}$$

Since $\tilde{T}-1\geq \frac{\tilde{T}}{2}$, we can derive the following sequence of inequalities:

$$\begin{aligned}
e^{-\vartheta}T^{-C\vartheta}(\tilde{T}-1) &\geq e^{-\vartheta}T^{-C\vartheta}\frac{\tilde{T}}{2} \\
&= e^{-\vartheta}T^{-C\vartheta}\frac{T}{2\Delta} \geq e^{-\vartheta}T^{1-C\vartheta}(2\lceil C\log T\rceil)^{-1} \geq B\left(\vartheta,T\right) \geq \frac{1}{\beta^2},
\end{aligned}$$

which hold as $T\geq n_1\geq n_0$.

Consequently, we obtain the following inequalities:

$$\mathbb{P}^{\pi}_{\ell_{\tilde{T}}}\left[|\hat{k}-\tau| > \Delta/3\right] \geq \frac{1}{\beta^2} \min_{1\leq j\leq \tilde{T}-1} \mathbb{P}^{\pi}_{\ell_j}\left[|\hat{k}-\ell_j| \leq \Delta/3\right] \geq \frac{\beta}{\beta^2} = \frac{1}{\beta} > 1,$$

which is clearly a contradiction. Therefore, we conclude:

$$\min_{1\leq j\leq \tilde{T}-1} \mathbb{P}^{\pi}_{\ell_j}\left[|\hat{k}-\tau| > \Delta/3\right] \geq 1-\beta.$$

That is, for all $T \geq n_1$, we obtain the following inequality:

$$\max_{1\leq \tau\leq T} \mathbb{P}^{\pi}_{\tau}\left[|\hat{k}-\tau| > \lceil C\log T\rceil/3\right] \geq 1-\beta,$$

where $C = \frac{1}{2\vartheta}$. That is, both n_1, as well as C only depends on parameters ϑ.

Case 2. In the second case, we assume that $T < n_1$. Then, we define $\tau_1 = T-1$ and $\tau_2 = T$. Suppose, first, that $\mathbb{P}_{\tau_1}\left[|\hat{k}-\tau_1| = 0\right] \geq \frac{1}{2}$. Next, observe that:

$$\begin{aligned}\mathbb{P}^{\pi}_{\tau_2}\left[|\hat{k}-\tau_1| = 0\right] &= \mathbb{E}^{\pi}_{\tau_2}\left[\mathbf{1}\big(|\hat{k}-\tau_1| = 0\big)\right] \\ &\overset{(a)}{=} \mathbb{E}^{\pi}_{\tau_1}\left[\exp\left(\log\Big(\frac{F^1\big(Z^{T-1} \mid \psi_{T-1}(\mathcal{H}_{T-2})\big)}{F^{\tau}\big(Z^{T-1} \mid \psi_{T-1}(\mathcal{H}_{T-2})\big)}\Big)\right)\mathbf{1}\big(|\hat{k}-\tau_1| = 0\big)\right],\end{aligned}$$

where (a) is obtained through change of measure type of argument, similar to the one employed within the Case 1. Hence, we have that:

$$\mathbb{P}^{\pi}_{\tau_2}\left[|\hat{k}-\tau_1| = 0\right] \geq e^{-\vartheta}\mathbb{E}_{\mathbb{P}^{\pi}_{\tau_1}}\left[\mathbf{1}\big(|\hat{k}-\tau_1| = 0\big)\right] \geq e^{-\vartheta}\frac{1}{2}.$$

Therefore, we obtain the following inequalities:

$$\mathbb{P}^{\pi}_{\tau_2}\left[|\hat{k}-\tau_2| \geq 1\right] = \mathbb{P}^{\pi}_{\tau_2}\left[|\hat{k}-T| \geq 1\right] \geq \mathbb{P}^{\pi}_{\tau_2}\left[|\hat{k}-\big(T-1\big)| = 0\right] \geq \frac{1}{2}e^{-\vartheta}.$$

On the other hand, if $\mathbb{P}_{\tau_1}\left[|\hat{k}-\tau_1| = 0\right] < \frac{1}{2}$, then we have:

$$\mathbb{P}_{\tau_1}\left[|\hat{k}-\tau_1| \geq 1\right] > \frac{1}{2} \geq \frac{1}{2}e^{-\vartheta}.$$

Therefore, we obtain the following inequality:

$$\sup_{\tau\in\{\tau_1,\tau_2\}} \mathbb{P}^{\pi}_{\tau}\left[|\hat{k}-\tau| \geq 1\right] \geq \frac{1}{2}e^{-\vartheta}.$$

Since $1 \geq T/n_1 \geq \log(T)/n_1$, we have that:

$$\sup_{1\leq\tau\leq T} \mathbb{P}^{\pi}_{\tau}\left[|\hat{k}-\tau| \geq \log(T)/n_1\right] \geq \frac{1}{2}e^{-\vartheta}.$$

Finally, by combining both Case 1 and Case 2, we obtain the following result:

$$\sup_{1\leq\tau\leq T} \mathbb{P}^{\pi}_{\tau}\left[|\hat{k}-\tau| \geq C_1\log T\right] \geq \frac{e^{-\vartheta}}{2} \geq \alpha,$$

where $\tilde{C} := \min\left\{C/3, 1/n_1\right\}$ and $\alpha := \min\left\{1-\beta, e^{-\vartheta}/2\right\}$. Moreover, both $\tilde{C} \equiv \tilde{C}(\vartheta)$ and $\alpha \equiv \alpha(\vartheta)$ only depends on ϑ, which concludes the proof. ■

E.C.4 Preliminaries

This section presents the notations and preliminary results that underpin the theoretical analysis in Section 5 and Appendix A.3. We provide formal statements and proofs of the key lemmas and corollary highlighted in Table E.C.1.

Lemma	**Corollary**
5, 6, 7, 8, 9	2

Table E.C.1: List of results from Appendix E.C.4

Throughout this section, we consider preferences $F^{(\mathbb{N})} \in \mathcal{F}_A$ in which a single abrupt shift occurs, as defined in Section 5.1. We refer to the pre-change preferences as F^1 and the post-change preferences as F^τ. Unless otherwise specified, we assume that both F^1 and F^τ are chosen such that, for some $\tau \in \mathbb{N}$, the preferences $F^{(\mathbb{N})} \equiv \big(F^t \,:\, t \in \mathbb{N}\big)$ satisfy $F^t \equiv F^1$ for all $t < \tau$, and $F^t \equiv F^\tau$ for all $t \geq \tau$, with the additional condition that $F^{(\mathbb{N})} \in \mathcal{F}_A$ (or a specified subset thereof). The variable $\tau \in [T+1]$ denotes the time at which the change occurs, where $\tau = T+1$ indicates that no change takes place. Finally, we denote by $\mathbb{P}^\pi_\tau$ the distribution over purchase outcomes induced by policy π when the change occurs at time τ, evaluated over the finite horizon T. Moreover, unless stated otherwise, we use notations consistent with those introduced in Section E.C.1.

Next, we introduce the *maximum revenue separation*:

$$\delta \equiv \delta(\mathcal{F}_A, T) := \sup\big\{r(\tilde{S}, F^t) - r(S, F^t) :\, F^{(\mathbb{N})} \in \mathcal{F}_A\,,\, t \in \mathbb{N}\,,\, S \neq \tilde{S} \in \mathcal{S}\big\} \leq N \cdot \|\mathbf{w}\|_\infty,$$

which captures the highest difference in expected revenue between any two distinct assortments.

To formally introduce our analysis, we begin by defining the random variable $\mathcal{J}^\pi(F^{(\mathbb{N})}, T)$, which represents the expected profit of a given policy π over a selling horizon of T periods. Specifically:

$$\mathcal{J}^\pi(F^{(\mathbb{N})}, T) := \sum_{t=1}^{\tau-1} \sum_{i \in \psi_t(\mathcal{H}_{t-1})} w_i p_i(\psi_t(\mathcal{H}_{t-1}), F^1) + \sum_{t=\tau}^{T} \sum_{i \in \psi_t(\mathcal{H}_{t-1})} w_i p_i(\psi_t(\mathcal{H}_{t-1}), F^\tau),$$

where the assortment policy is $\pi \in \mathcal{P}$, defined as $\pi := (\psi_t(\mathcal{H}_{t-1}) \;:\; 1 \leq t \leq T)$. Next, we show that the expected value of the random variable $\mathcal{J}^\pi(F^{(\mathbb{N})}, T)$, taken with respect to $\mathbb{P}^\pi_\tau$, is simply the expected cumulative revenue achieved by policy π.

Lemma 5. *For $T \geq 2$, we have that $J^\pi(F^{(\mathbb{N})}, T) = \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\mathcal{J}^\pi(F^{(\mathbb{N})}, T)\big]$.*

Proof. In the following proof, we omit the dependence of the policy $\pi \in \mathcal{P}$ on the filtration $(\mathcal{H}_t)_{t=0}^T$

to simplify the notations. Specifically, we denote $\psi_t\left(\mathcal{H}_{t-1}\right)$ simply by ψ_t for $t \in [T]$. Then:

$$
\begin{aligned}
J^\pi(F^{(\mathbb{N})},T) &= \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{T}\sum_{i\in\psi_t} w_i\mathbf{1}(i_t=i)\big] = \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\tau-1}\sum_{i\in\psi_t} w_i\mathbf{1}(i_t=i)\big] + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i\mathbf{1}(i_t=i)\big],\\
&= \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\tau-1}\sum_{i\in\psi_t} w_i\mathbf{1}\big(i_t=i\in\psi_t, U^a_{i_t}>U^a_j,\ \forall\, j\in\psi_t\cup\{0\}\setminus\{i_t\}\big)\big]\\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i\mathbf{1}\big(i_t=i\in\psi_t, U^b_{i_t}>U^b_j,\ \forall\, j\in\psi_t\cup\{0\}\setminus\{i_t\}\big),\big]\\
&\overset{(a)}{=} \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\tau-1}\sum_{i_t=i\in\psi_t} w_i\mathbf{1}\big(i_t=i\in\psi_t, U^a_{i_t}>U^a_j,\ \forall\, j\in\psi_t\cup\{0\}\setminus\{i_t\}\big)\mid\mathcal{H}_{t-1}\big]\big]\\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i\mathbf{1}\big(i_t=i\in\psi_t, U^b_{i_t}>U^b_j,\ \forall\, j\in\psi_t\cup\{0\}\setminus\{i_t\}\big)\mid\mathcal{H}_{t-1}\big]\big]\\
&\overset{(b)}{=} \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\tau-1}\sum_{i\in\psi_t} w_i\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\mathbf{1}\big(i_t=i\in\psi_t, U^a_{i_t}>U^a_j,\ \forall\, j\in\psi_t\cup\{0\}\setminus\{i_t\}\big)\mid\mathcal{H}_{t-1}\big]\big]\\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\mathbf{1}\big(i_t=i\in\psi_t, U^b_{i_t}>U^b_j,\ \forall\, j\in\psi_t\cup\{0\}\setminus\{i_t\}\big)\mid\mathcal{H}_{t-1}\big]\big]\\
&\overset{(c)}{=} \sum_{t=1}^{\tau-1}\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{i\in\psi_t} w_i p_i\left(\psi_t,F^1\right)\big] + \sum_{t=\tau}^{T}\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{i\in\psi_t} w_i p_i\left(\psi_t,F^\tau\right)\big] = \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\mathcal{J}^\pi(F^{(\mathbb{N})},T)\big],
\end{aligned}
$$

where, step (a) follows from the Law of Total Expectation (Jacod and Protter 2012), step (b) follows from moving the summation outside the expectation, and step (c) is a consequence of the definition of the probability of purchase $p_i\left(S,F\right)$. ■

As a consequence of Lemma 5, the difference between the expected revenue achieved by the oracle and that achieved by policy π can be decomposed into two components: the regret incurred before the change occurs, and the regret incurred after the change. Formally,

Corollary 2. *For $F^{(\mathbb{N})}\in\mathcal{F}_A$, we have that, for $T\geq 2$:*

$$
\begin{aligned}
J^*(F^{(\mathbb{N})},T) - J^\pi(F^{(\mathbb{N})},T) &= \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=1}^{\tau-1}\big(r(S^*(F^1),F^1) - r(\psi_t(\mathcal{H}_{t-1}),F^1)\big)\big]\\
&\quad + \mathbb{E}_{\mathbb{P}^\pi_\tau}\big[\sum_{t=\tau}^{T}\big(r(S^*(F^\tau),F^\tau) - r(\psi_t(\mathcal{H}_{t-1}),F^\tau)\big)\big].
\end{aligned}
$$

Proof. The proof immediately follows from Lemma 5. ■

Next, we formally define and derive a closed-form expression for the probability of purchase over a finite sequence of T customers, denoted by $\mathbb{P}^\pi_\ell$. Customer purchase decisions are modeled as a random vector Z with dimension $T \times N$. Moreover, for any assortment $S \in \mathcal{S}$, we denote by $F(\cdot \mid S)$ the conditional distribution of F given the offered assortment S.

Lemma 6. *Let $z \in \{0,1\}^{T\times N}$ be some customers' purchase realization, and $\pi \in \mathcal{P}$ an admissible policy such that $\pi := (\psi_t(\mathcal{H}_{t-1}) \mid 1 \le t \le T)$. Then, for any given time $\ell \in [T]$, we have:*

$$\mathbb{P}_\ell^\pi\big[Z = z\big] = \prod_{t=1}^{\ell-1} F^1(z^t \mid \psi_t(\mathcal{H}_{t-1})) \prod_{t=\ell}^{T} F^\tau(z^t \mid \psi_t(\mathcal{H}_{t-1})).$$

In particular, if there exists $t \in [T]$ such that $z_i^t = 1$ for some $i \in \mathcal{N}$ and $i \notin \psi_t(\mathcal{H}_{t-1})$, then:

$$\mathbb{P}_\ell^\pi\big[Z = z\big] = 0.$$

Proof. All customers are assumed to act independently according to their own intrinsic utility. Accordingly, the distribution of the purchase decision of customer t is independent of the purchase decisions of customers 1 to $t-1$. Thus, the following sequence of equalities holds:

$$\begin{aligned}
\mathbb{P}_\ell^\pi\big[Z = z\big] &= \mathbb{P}_\ell^\pi\big[Z^t = z^t,\ 1 \le t \le T\big] \\
&= \prod_{t=1}^{T} \mathbb{P}_\ell^\pi\big[Z^t = z^t \mid Z^{\tilde{t}} = z^{\tilde{t}},\ 1 \le \tilde{t} \le t-1\big] \\
&= \prod_{t=1}^{T} \mathbb{P}_\ell^\pi\big[Z^t = z^t \mid \mathcal{H}_{t-1}\big] \\
&= \prod_{t=1}^{\ell-1} F^1(z^t \mid \pi\ , \mathcal{H}_{t-1}) \prod_{t=\ell}^{T} F^\tau(z^t \mid \pi\ , \mathcal{H}_{t-1}) = \prod_{t=1}^{\ell-1} F^1(z^t \mid \psi_t(\mathcal{H}_{t-1})) \prod_{t=\ell}^{T} F^\tau(z^t \mid \psi_t(\mathcal{H}_{t-1})).
\end{aligned}$$

Next, assume that there exists $t \in [T]$ such that $z_{i,t} = 1$ for some product $i \in \mathcal{N}$, which does not belong to the assortment ψ_t, that is, $\psi_t(\mathcal{H}_{t-1})_i = 0$. Then, recall that $p_i(\psi_t(\mathcal{H}_{t-1}), F) = 0$, for all $i \notin \psi_t$, and $F \in \{F^1, F^\tau\}$. Consequently, $F^1(z^t \mid \psi_t(\mathcal{H}_{t-1})) = 0$ if $t \le \ell - 1$, and $F^\tau(z^t \mid \psi_t(\mathcal{H}_{t-1})) = 0$ if $t \ge \ell$. Therefore, we have that $\mathbb{P}_\ell^\pi\big[Z = z\big] = 0$, which concludes the proof. ■

The distribution $\mathbb{P}_\ell^\pi$ denotes the probability measure induced over the purchase outcomes for a particular change scenario, parameterized by the change time. The similarity between two such scenarios is quantified using the KL divergence between their respective distributions. Lemma 7 provides a closed-form expression for this divergence.

Lemma 7. *For $\ell_j, \ell_{j+1} \in \{1, \dots, T\}$, let $\mathcal{K}(\mathbb{P}_{\ell_{j+1}}^\pi, \mathbb{P}_{\ell_j}^\pi)$ denote the KL divergence between the two probability measures $\mathbb{P}_{\ell_j}^\pi$ and $\mathbb{P}_{\ell_{j+1}}^\pi$. Then, we have that:*

$$\mathcal{K}(\mathbb{P}_{\ell_{j+1}}^\pi, \mathbb{P}_{\ell_j}^\pi) = \sum_{t=\ell_j}^{\ell_{j+1}-1} \mathbb{E}_{\mathbb{P}_{\ell_{j+1}}^\pi}\big[\log \frac{F^1(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}{F^\tau(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}\big].$$

Proof. To begin, the KL divergence between $\mathbb{P}^{\pi}_{\ell_{j+1}}$ and $\mathbb{P}^{\pi}_{\ell_j}$ is formally defined as follows:

$$\mathcal{K}(\mathbb{P}^{\pi}_{\ell_{j+1}}, \mathbb{P}^{\pi}_{\ell_j}) = \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{j+1}}}\big[\log \frac{\mathbb{P}^{\pi}_{\ell_{j+1}}\big[Z\big]}{\mathbb{P}^{\pi}_{\ell_j}\big[Z\big]}\big],$$

where Z is the random vector representing the customer's purchase decisions over the horizon T.

Next, we leverage the closed-form formula for the distribution of Z from Lemma 6 to obtain the desired result. Specifically, the following sequence of equalities holds:

$$\begin{aligned}\mathcal{K}(\mathbb{P}^{\pi}_{\ell_{j+1}}, \mathbb{P}^{\pi}_{\ell_j}) &= \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{j+1}}}\left[\log \frac{\prod_{t=1}^{\ell_{j+1}-1} F^1(Z^t \mid \psi_t(\mathcal{H}_{t-1})) \prod_{t=\ell_{j+1}}^{T} F^{\tau}(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}{\prod_{t=1}^{\ell_j-1} F^1(Z^t \mid \psi_t(\mathcal{H}_{t-1})) \prod_{t=\ell_j}^{T} F^{\tau}(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}\right] \\ &= \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{j+1}}}\left[\log \frac{\prod_{t=\ell_j}^{\ell_{j+1}-1} F^1(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}{\prod_{t=\ell_j}^{\ell_{j+1}-1} F^{\tau}(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}\right] = \sum_{t=\ell_j}^{\ell_{j+1}-1} \mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{j+1}}}\big[\log \frac{F^1(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}{F^{\tau}(Z^t \mid \psi_t(\mathcal{H}_{t-1}))}\big],\end{aligned}$$

which concludes the proof. ■

In the following lemma, we define the maximum KL divergence between F^1 and F^{τ}, conditional on an assortment $S \in \mathcal{S}$. By the definition of $\mathcal{F}$ (in Section 3), we have $F(z \mid S) \in (0, 1)$ for all $z \in \{0, 1\}^N$ such that $\|z\|_1 \leq 1$, where $z_i = 0$ for any $i \notin S$, given $S \in \mathcal{S}$ and $F \in \{F^1, F^{\tau}\}$. Since $F^1(z \mid S) = 0$ whenever $z_i = 1$ for some $i \notin S$, the KL divergence between $F^1(\cdot \mid S)$ and $F^{\tau}(\cdot \mid S)$ is well-defined, ensuring $\mathcal{K}(F^1\,,\,F^{\tau}\,;\,S) < \infty$. Furthermore, because the set of assortments $\mathcal{S}$ is finite, the maximum KL divergence (taken over all assortments $S \in \mathcal{S}$) is also well-defined. The maximum KL divergence between F^1 and F^{τ}, conditional on S, is defined by :

$$\mathcal{K}(F^1, F^{\tau}) \equiv \max\big\{\mathcal{K}(F^1\,,\,F^{\tau}\,;\,S)\;:\;S \in \mathcal{S}\big\}.$$

Lemma 8. *We have that* $0 < \mathcal{K}(F^1, F^{\tau}) < \infty$.

Proof. By the definition of KL divergence, together with the definition of $\mathcal{F}$, we have that, for any $S \in \mathcal{S}$, $0 \leq \mathcal{K}(F^1\,,\,F^{\tau}\,;\,S) < \infty$. Since $\mathcal{S}$ is finite, it follows that:

$$0 \leq \mathcal{K}(F^1, F^{\tau}) < \infty.$$

We show that $\mathcal{K}(F^1, F^{\tau}) > 0$ by contradiction. Hence, assume for the sake of contradiction that $\mathcal{K}(F^1\,,\,F^{\tau}\,;\,S) = 0$ for all $S \in \mathcal{S}$. By the properties of the KL divergence, this implies $F^1(\cdot \mid S) = F^{\tau}(\cdot \mid S)$ for all $S \in \mathcal{S}$. Consequently, the optimal assortments $S^*(F^1)$ and $S^*(F^{\tau})$, corresponding to F^1 and F^{τ}, respectively, must be identical, i.e., $S^*(F^1) = S^*(F^{\tau})$. This result contradicts the assumption that the pre- and post-change optimal assortment are different. Therefore, we conclude that $\mathcal{K}(F^1, F^\tau) > 0$, completing the proof. ■

Next, we assume that $T \geq 2$ is fixed and segment the time horizon T into sub-segments of size $\Delta \in [T]$. Let $\tilde{T} - 1 = \lceil T/\Delta \rceil - 1$ denote the number of such sub-segments. We define the indices $(\ell_j)_{j=0}^{\tilde{T}-1}$ as follows: $\ell_0 = 1$, and $\ell_j = \ell_{j-1} + \Delta$ for $j \in [\tilde{T} - 1]$. Note that the final sub-segment, $\tilde{T} - 1$, may have a cardinality less than Δ. Moreover, given two assortments $S, \tilde{S} \in \mathcal{S}$, we denote by $\|S - \tilde{S}\|_1$ the Hamming distance between their respective binary encodings.

Lemma 9. *Assume that the two distributions F^1 and F^τ are equal conditional on the assortment $S^*(F^1)$. Then, given some index $j \in [\tilde{T} - 1]$, the following inequality holds:*

$$\mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\sum_{t=\ell_j}^{\ell_{j+1}-1} \mathbf{1}\big(\|\psi_t(\mathcal{H}_{t-1}) - S^*(F^1)\|_1 > 0\big)\big] \geq \frac{\mathcal{K}(\mathbb{P}^\pi_{\ell_{j+1}}, \mathbb{P}^\pi_{\ell_j})}{\mathcal{K}(F^1, F^\tau)}.$$

Proof. To simplify the notations within this proof, we omit the explicit dependence of π on the filtration $(\mathcal{H}_t)_{t=1}^T$. That is, we refer to $\psi_t\left(\mathcal{H}_{t-1}\right)$ as ψ_t for all $t \in \left[T\right]$. By Lemma 7, we have

$$\mathcal{K}(\mathbb{P}^\pi_{\ell_{j+1}}, \mathbb{P}^\pi_{\ell_j}) = \sum_{t=\ell_j}^{\ell_{j+1}-1} \mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\log \frac{F^1\left(Z^t \mid \psi_t\right)}{F^\tau(Z^t \mid \psi_t)}\big].$$

For $t \in \{\ell_j, \ldots, \ell_{j+1} - 1\}$, by using the Law of Total Expectation (Jacod and Protter 2012), we obtain the following equality:

$$\mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\log \frac{F^1(Z^t \mid \psi_t)}{F^\tau(Z^t \mid \psi_t)}\big] = \mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\log \frac{F^1(Z^t \mid \psi_t)}{F^\tau(Z^t \mid \psi_t)} \mid \psi_t\big]\big].$$

Fix a feasible assortment $S \in \mathcal{S}$ arbitrarily. If the purchase decisions over the all the time periods are $\mathbb{P}^\pi_{\ell_{j+1}}$ distributed, then the random vector Z^t, which models consumer purchase decision at time t, is F^1 distributed (conditional on assortment ψ_t). Thus, the following equalities hold:

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\log \frac{F^1(Z^t \mid \psi_t)}{F^\tau(Z^t \mid \psi_t)} \mid \psi_t(\mathcal{H}_{t-1}) = S\big] &= \sum_{z \in \{0.1\}^N} \mathbf{1}(z_i = 0\,, \forall\, i \notin S) F^1(z \mid S) \log \frac{F^1(z \mid S)}{F^\tau(z \mid S)} \\ &= \mathcal{K}(F^1\,,\, F^\tau\,;\, S).\end{aligned}$$

Therefore, the following sequence of inequalities holds:

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\log \frac{F^1\left(Z^t \mid \psi_t\right)}{F^\tau\left(Z^t \mid \psi_t\right)}\big] &= \mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\mathcal{K}(F^1\,,\, F^\tau\,;\, \psi_t)\mathbf{1}\big(\psi_t = S^*(F^1)\big)\big] \\ &\quad + \mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\mathcal{K}(F^1\,,\, F^\tau\,;\, \psi_t)\mathbf{1}\big(\psi_t \neq S^*(F^1)\big)\big] \\ &\overset{(a)}{=} \mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\mathcal{K}(F^1\,,\, F^\tau\,;\, \psi_t)\mathbf{1}\big(\psi_t \neq S^*(F^1)\big)\big] \\ &\overset{(b)}{\leq} \mathcal{K}(F^1, F^\tau)\mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[\mathbf{1}\big(\|\psi_t - S^*(F^1)\|_1 > 0\big)\big],\end{aligned}$$

where (a) follows by the assumption that F^1 and F^τ are different, conditional on the pre-change optimal assortment $S^*(F^1)$, and (b) follows by the definition of $\mathcal{K}(F^1, F^\tau)$.

Therefore, summing over t yields:

$$\mathcal{K}(\mathbb{P}^\pi_{\ell_{j+1}}, \mathbb{P}^\pi_{\ell_j}) \leq \mathcal{K}(F^1, F^\tau) \sum_{t=\ell_j}^{\ell_{j+1}-1} \mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}} \left[\mathbf{1}\left(\|\psi_t - S^*(F^1)\|_1 > 0 \right) \right],$$

which concludes the proof. ■

References

Besbes, O. and Zeevi, A. (2011). "On the minimax complexity of pricing in a changing environment". In: *Operations Research* 59.1, pp. 66–79.

Jacod, J. and Protter, P. (2012). *Probability essentials*. Springer Science & Business Media.

Korostelev, A. (1988). "On minimax estimation of a discontinuous signal". In: *Theory of Probability & Its Applications* 32.4, pp. 727–730.

Naaman, M. (2021). "On the tight constant in the multivariate Dvoretzky–Kiefer–Wolfowitz inequality". In: *Statistics & Probability Letters* 173, p. 109088.

Thomas, M. and Joy, A. T. (2006). *Elements of information theory*. Wiley-Interscience.

Tsybakov, A. B. (2003). *Introduction à l'estimation non paramétrique*. Vol. 41. Springer Science & Business Media.