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

May 6, 2025

Abstract. We examine a dynamic assortment planning problem in which a retailer operates under evolving and unobservable customers' preferences and seeks to maximize revenue. These evolving preferences are characterized along three key dimensions—*velocity*, *magnitude*, and *detectability*—that, as we show, jointly shape the complexity of the assortment planning task. We assess performance via the revenue gap relative to a clairvoyant retailer with full knowledge of preferences and design policies that exploit the characteristics of the preferences dynamics to "learn" from customers' purchasing decisions. Our findings highlight a paradox: slow but small changes in preferences can lead to substantial performance losses, akin to the "boiling frog" apologue, as failure to respond to gradual market changes results in missed opportunities. We show that, to mitigate this risk efficiently, the retailer should frequently re-estimate up-to-date preferences to avoid offering an outdated assortment. Furthermore, we explore how information about preferences' characteristics enables the design of improved assortment strategies. In particular, we consider scenarios in which the retailer anticipates an abrupt change in preferences. In such cases, frequently re-estimating preferences provides a practical hedge against this market dynamic. Yet, when more information on preferences is available to the retailer, tailored strategies based on detecting a change in preferences substantially improve performance. Through theoretical analysis and empirical validation using data from a major Chilean retailer, we demonstrate the value of aligning assortment strategies with the dynamic nature of preferences and the available information on them.

1 Introduction

Since the advent of e-commerce in the late 1990s and concurrent advances in information technology, online retailers have benefited from unprecedented opportunities to display and update their product mixes with minimal friction (Caro et al. 2020). Not only do online channels enable personalized targeting (Bernstein et al. 2019), but they also deliver unparalleled insights into consumer purchasing patterns. However, despite all these advantages, online platforms face a critical

constraint: the limited display space on key pages (e.g., homepages and search results). This limitation forces them to be highly selective about the products they showcase. Consequently, they must carefully align their assortments with customers' preferences to maximize revenue from sales.

The dynamic nature of customers' preferences has been recognized since the 1960s with early insights from behavioral learning theory shaping our understanding of brand loyalty (J. N. Sheth 1967). Recognizing the impact of evolving customers' preferences offers an opportunity to better understand the market in which retailers operate (Hartmann and Nair 2010). Whether operating in brick-and-mortar stores or online, retailers should consider changing consumer behavior when setting up core operational functions, ranging from pricing to assortment planning (Lattin 1987; Chintagunta et al. 2002). Nevertheless, a large portion of the assortment planning literature still assumes that customers' preferences remain static, or, put in another way, time-homogeneous.

Demographic trends and the slow evolution of long-term tastes can subtly yet persistently modify consumer behavior (Döpper et al. 2024)—a dynamic that becomes particularly salient when viewed through the lens of assortment planning. Indeed, due to its combinatorial nature, small changes in preferences might modify drastically the assortment decision. Moreover, abrupt disruptions—such as pandemics, financial crises, or viral social media phenomena—can rapidly overturn established purchasing patterns, forcing retailers to quickly adapt their assortment offering accordingly. For instance, recessions have been shown to drive consumers toward budget-friendly alternatives, disadvantaging high-end products (Hampson and McGoldrick 2013). Likewise, pandemic outbreaks can cause abrupt surges in demand for essential products like medical supplies, while simultaneously depressing sales of services due to the sanitary restrictions in place (J. Sheth 2020).

Accordingly, a retailer's ability to recognize and respond to evolving customers' preferences can create opportunities for growth or, conversely, leave it vulnerable to dynamic market forces. To illustrate the stakes, consider the boiling frog apologue[1]: when a frog is placed in water that is gradually heated, it fails to perceive gradually increasing danger until it becomes too late. Similarly, a retailer that fails to recognize (or misreads) the persistent evolution of customers' preferences risks being caught off guard. However, accurately identifying these dynamics is not always straightforward, as some changes may go unnoticed by a retailer following a pre-established assortment plan.

Objective. We address the problem of dynamic assortment planning for a retailer facing customers whose preferences evolve over time. Our aim is to characterize the complexity of the assortment planning problem for retailers in terms of the information they have on the evolving preferences. Accordingly, we characterize changes along three dimensions: (*i*) the *velocity*, (*ii*) the *mag-*

nitude, and (*iii*) the *detectability*. When changes occur within the assortment offered by the retailer, we say that they can be detected *passively* by monitoring sales data. In contrast, changes involving products outside the current offering require *active* exploration through alternative assortments.

Model. We consider dynamic assortment planning over a known horizon T, where customers arrive sequentially and decide whether to purchase from the offered assortment. The retailer aims to maximize profit but operates with uncertainty regarding how customers' preferences evolve over time. Our problem lies within the broader domain of sequential decision-making under uncertainty (Hannan 1957). To address the challenges posed by evolving preferences, we propose assortment strategies that adapt "on the fly" to changes in purchasing behavior.

We evaluate the retailer's performance by comparing the expected revenue it can achieve under a given *policy* (an assortment strategy) to that of a clairvoyant retailer (an oracle) endowed with perfect knowledge of preferences. This difference is commonly referred to as *regret*, a well-established metric in the online learning literature (Foster and Vohra 1999). It serves as a proxy to capture the *opportunity cost* from not having complete information on customers' preferences. In particular, we focus on the worst-case regret and its dependence on the time horizon T. This measure can be viewed through a game-theoretic lens: the retailer selects an assortment strategy first, and the environment then responds adversarially by determining customers' preferences.

Under static preferences, dynamic assortment planning typically tackles the classical trade-off between *exploration* and *exploitation* (Lai and Robbins 1985), balancing the need to explore different assortments to collect data on consumer purchasing behavior against the goal of exploiting current estimates to maximize immediate revenue (Caro and Gallien 2007). These algorithms typically rely on an initial exploration step followed by an exploitation period (Sauré and Zeevi 2013). We argue that this approach is inadequate in the case of evolving customers' preferences, "persistent exploration" becomes essential in our setting. In particular, we address the trade-off between exploration and exploitation, and examine how it is influenced by the structural information available to the retailer. Our approach is modular, meaning that any improvement in dynamic assortment planning for static preferences directly enhances the performances of our proposed methods.

Technical contributions. We bridge diverse strands of research, ranging from dynamic pricing to stochastic programming, to present a unified characterization of dynamic assortment planning with evolving customers' preferences. Our contributions, summarized in Table 1, are threefold.

First, we analyze the case in which the retailer is equipped only with information on the "magnitude" of changes in customers' preferences. This quantity, denoted by M_T, provides a clear

Velocity	Magnitude	Detectability	Preferences	Regret (informal)	Section
-	M_T	-	Unknown	$\mathcal{O}\big(\mathcal{R}^{\mathcal{A}}\big(T^{\frac{1}{2}}M_T^{-\frac{1}{2}}\big)\cdot T^{\frac{1}{2}}M_T^{\frac{1}{2}}\big)$	4
Abrupt	M_T	Active	Unknown	$\mathcal{O}\big(\mathcal{R}^{\mathcal{A}}(TM_T)\cdot M_T^{-1}\big)$	5.2
Abrupt	bounded	Active	Unknown	$\mathcal{O}\big(\sqrt{T}\log T+\mathcal{R}^{\mathcal{A}}(T)\big)$	5.3
Abrupt	$\geq$ constant	Passive	Unknown	$\mathcal{O}\big(\log T+\mathcal{R}^{\mathcal{A}}(T)\big)$	5.4
Abrupt	-	Active	Known	$\mathcal{O}\big(\sqrt{T}\log T\big)$	A.1
Abrupt	-	Passive	Known	$\mathcal{O}\big(\log T\big)$	A.2

Table 1: Summary of our contributions. We categorize preferences changes along three dimensions: velocity, magnitude, and detectability. The column "preferences" indicates whether the preferences after the change are known by the retailer. For each setting, we report the regret of the proposed assortment strategies, which depends on the information available to the retailer. Here, $\mathcal{R}^{\mathcal{A}}(T)$ denotes the regret of an assortment strategy $\mathcal{A}$ designed for static preferences over T periods, and M_T refers to the largest magnitude of change (formally defined later).

indication of how unstable or dynamic these preferences may be. In particular, we show that no assortment strategy can achieve a regret lower than $\mathcal{O}\big(T^{3/4}M_T^{1/4}\big)$. Although our proof techniques build upon the approach by Besbes et al. (2015), our bound deviates from theirs due to key differences in modeling variability of the environment. Notably, we reveal a jump in the opportunity cost compared to settings with time-homogeneous preferences in which assortment strategies typically achieve adversarial regret of order $\mathcal{O}\big(\sqrt{T\log T}\big)$; see Agrawal et al. (2019).

To match our lower bound, we propose a "restart-and-learn" algorithm that partitions the selling horizon into segments and applies a learning subroutine tailored to static customers' preferences within each segment. If this subroutine achieves regret on the order of $\mathcal{O}(\sqrt{T})$, then the overall policy matches the lower bound. In other words, the retailer can attain the best possible performance even when it has only limited information about the dynamics of customers' preferences.

Second, we consider the case in which the retailer anticipates a single abrupt shift in consumer behavior, transitioning from known *pre-change* preferences to unknown *post-change* preferences, with neither the timing nor the extent known a priori. Our analysis builds on Besbes and Zeevi (2011), who study a related setting in dynamic pricing. However, while their analysis assumes full knowledge of post-change demand and leverages specific structural conditions, we extend their approach to assortment planning under weaker informational and structural assumptions.

When limited additional information is available on the magnitude of the change, we show that any assortment strategy must incur a regret of at least $\mathcal{O}\big(T^{1/2}M_T^{-1/2}\big)$. We show that the restart-and-learn policy attains regret of order $\mathcal{O}\big(\mathcal{R}^{\mathcal{A}}(T\cdot M_T)\cdot M_T^{-1}\big)$, where $\mathcal{R}^{\mathcal{A}}(\Delta)$ represents the regret achieved by the assortment strategy $\mathcal{A}$ when serving Δ customers with preferences that remain static. These results indicate that the difficulty of the setting might not only arise from large and sudden changes but also from small and incremental ones.

Third, we discuss settings in which the retailer anticipates that preferences vary within a known range. We focus on the detectability of such changes, distinguishing two cases. A change is *passively detectable* if it affects products that would be offered under a strategy specifically designed for static preferences. On the other hand, the change is said to be *actively detectable* if it only affects products that would not be offered under such strategy, thus requiring "active" exploration.

For actively detectable changes, we establish a regret lower bound of order $\mathcal{O}(\sqrt{T})$. We then propose an active assortment strategy where the retailer strategically deploys multiple assortments and applies a change detection procedure to identify shifts in preferences. This strategy achieves a regret of order $\mathcal{O}\big(\sqrt{T}\log T\big)$, in addition to that from learning the new preferences.

Then, for passively detectable changes, our approach follows a strategy designed for (known) static preferences and applies a change detection procedure with a certain frequency. Once a change is detected, the retailer transitions to a learning algorithm to adapt to new preferences, incurring a regret of order $\mathcal{O}\big(\log T\big)$ on top of that from learning the new preferences. Notably, in this setting, exploration and exploitation are not in conflict and can be pursued somewhat simultaneously.

Managerial insights. Evolving customers' preferences pose a hidden threat to retailers' assortment strategies—much like how gradually rising water temperature can fatally catch a frog off guard. When no additional information on the changes is available, continuously restarting the learning process helps avoid being "boiled." On the other hand, our result shows that when structural information on the velocity, magnitude, and detectability is gathered by the retailer, then it can be exploited to improve performance. Notably, when preferences are expected to change abruptly by a minimal amount, effective policies focus on change detection rather than periodic resets. Finally, a case study using data from a large Chilean retailer confirms the practical benefits of leveraging structural information on the change and underscores the necessity of staying responsive to evolving preferences.

Organization of the paper. Section 2 reviews the relevant literature. In Section 3, we provide a formulation of the dynamic assortment planning problem with evolving customers' preferences. Section 4 presents an algorithm to handle these variations and establishes performance guarantees that no algorithm can surpass. Next, Section 5 addresses abrupt changes in purchase pattern, where we establish fundamental limits on policy performance and propose effective assortment strategies. In Section 6, we present a case study with click data from a major Chilean retailer. Finally, Section 7 provides concluding remarks. All proofs are provided in the electronic companion.

Notations. For $n \in \mathbb{N}$, let $[n] \coloneqq \{1, \ldots, n\}$, and denote the cardinality of a set A by $|A|$. All

random variables are defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$.

2 Literature review

Consumer 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). Choice models typically represent customers' preferences by linking the utility they attach to each product to purchase probabilities (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, complicating its estimation. To address the computational burden that these models pose in assortment planning, approximation algorithms are often derived (Golrezaei et al. 2014; Blanchet et al. 2016). In parallel, non-parametric approaches—typically based on consumers' product rankings—have also attracted interest (Honhon et al. 2012; Van Ryzin and Vulcano 2015; Bertsimas and Mišić 2019).

Role of learning. Since customers' preferences are typically unknown to the retailer, studies have emerged on recovering 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 structure directly from sales data. In the *online* setting, the retailer learns by continuously updating its estimate of preferences "on the fly" from both the observed sales data and the assortments displayed to the customers.

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).

Learning in varying environments. In dynamic retail environments, customers' preferences might not always be time-homogeneous, yet most traditional assortment models assume otherwise. This "non-stationarity" is typically addressed by allowing demand parameters to vary over time. For example, Golrezaei et al. (2014) discuss non-stationarity under the restrictive assumption that the retailer knows exactly how preferences change. Similarly, Foussoul et al. (2023) consider a multi-armed bandit problem which involve a time-varying MNL model. Though their model fits within our broader approach, our study takes a different direction, examining how structural information on customers' preferences may influence the retailer's assortment strategy and the resulting revenue.

In the broader context of revenue management, the challenge of learning the demand function in dynamic pricing is addressed by Besbes and Zeevi (2009), and is extended by Besbes and Zeevi (2011), Besbes and Sauré (2014), and Keskin and Zeevi (2017) to account for changing demand. Beyond revenue management, Besbes et al. (2015) consider a non-stationary stochastic optimization problem with unknown, time-varying cost functions constrained by a bounded variation budget. They establish a fundamental lower bound on the regret of $\mathcal{O}(T^{2/3}M_T^{1/3})$, where M_T denotes the total variation of the cost functions, and propose a policy with regret that matches this bound.

In non-stationary online optimization, two main paradigms prevail. In the first, parameters are allowed to vary over time, subject to a bound on their cumulative variation (Besbes et al. 2015); in the second one, only a finite number of abrupt changes are permitted. For the latter, some algorithms rely on sliding window techniques to focus learning on recent data (Garivier and Moulines 2011), while others employ change-detection methods to reset and re-learn model parameters upon detecting a change (Zhou et al. 2020). In this abrupt-change setting, no policy can achieve regret below the fundamental $\mathcal{O}(\sqrt{T})$ bound by Garivier and Moulines (2011). By contrast, classical stationary settings yield instance-dependent regret rates of $\mathcal{O}(\log T)$ or, in adversarial cases, $\mathcal{O}(\sqrt{T})$.

Change detection for abrupt shocks. In this study, we underscore the importance of detecting abrupt changes in customers' preferences for effective assortment planning. This challenge is closely related to the classical *quickest detection problem* (Shiryaev 1963), where methods such as the sequential likelihood ratio test (Wald and Wolfowitz 1948; Lorden 1971) aim to identify distributional changes rapidly while keeping false alarms in check. Typically, these techniques assume that the post-change distribution is known (Pollak 1985); when it is not, parametric models (Lai 1998) can help manage uncertainty around both the timing and nature of the change. Although these detection approaches have traditionally been applied in statistical process control (Korostelev 1988),

we leverage them to update our assortments quickly by recognizing when customers' preferences have changed. Our work adopts a frequentist perspective on change detection, while acknowledging that Bayesian frameworks (Tartakovsky and Veeravalli 2005) also present powerful alternatives.

3 Problem formulation

Model primitives and basic assumptions. 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 customers' preferences. 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, so that any customer may be offered any assortment in $\mathcal{S}$. Although these assumptions are admittedly restrictive, they are commonly adopted in the dynamic assortment planning literature to isolate the effect of learning customers' preferences; see, e.g., Sauré and Zeevi (2013), Agrawal et al. (2019), and Li et al. (2025). For studies that incorporate inventory constraints, we refer to Mahajan and Van Ryzin (2001), Chen et al. (2024), and Zhang et al. (2024).

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 from the offered assortment $i \in S^t \cup \{0\}$, 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:

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

To focus on the dynamics of assortment planning rather than the isolated single-sale problem, we assume that the optimal assortment, denoted by $S_t^* \equiv S^*(F^t)$, is uniquely determined.

Dynamics of preferences. Unlike traditional assortment planning models, we consider settings where customers' preferences, that are represented by $F^{(\mathbb{N})}$, evolve over time rather than remaining static, i.e., $F^t \equiv F$ for all $t \in \mathbb{N}$ and some fixed distribution F. Since these preferences are not directly observable, the sequence $F^{(\mathbb{N})}$ is unknown *a priori*, although partial information—such as early-horizon preferences or structure on the choice model—may be available. We encapsulate this information in the set $\mathcal{F}$, which comprises all possible preferences sequences the retailer might encounter. Notably, we assume that the retailer specifies $\mathcal{F}$ in advance by imposing structure on the customers' choice process, either via a parametric model or a ranking-based approach. 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.

The retailer may wish to incorporate insights into how customers' preferences evolve over time or capture specific behavioral effects within its modeling. For example, it might be particularly interested in the time-inconsistent behavior of its customers. Specifically, despite maintaining stable long-term preferences (Wood and Neal 2009), consumers occasionally deviate due to self-control limitations. Empirical evidence from Hoch and Loewenstein (1991) indicates that consumers sometimes make choices they later regret. To capture such behaviors, the set $\mathcal{F}$ can be restricted accordingly. This restriction, in turn, determines the maximum magnitude of the change in preferences, denoted by $\mathcal{M}(\mathcal{F}, T)$ and defined in Section 4. Alternatively, preferences may evolve gradually over longer periods. Empirical evidence suggests that the adoption of new communication technologies follows a slow diffusion process driven by network effects (Tucker 2008). This gradual evolution can also be integrated into $\mathcal{F}$, and, in turn, is reflected in the dependence of $\mathcal{M}(\mathcal{F}, T)$ on T.

Assortment planning under evolving preferences. Let $\mathcal{H}_t := \sigma\big((S^s, i_s) \,:\, s < t\big)$ denote the history of offered assortments and consumer purchases prior to customer $t \in [T]$. A sequence

of (random) assortments $\pi := \left(S_t^\pi \ : \ t \leq T\right)$ 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 $\left(\mathcal{H}^s \ : \ s < t\right)$ 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:

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

In essence, the oracle revenue represents the best achievable profit and is not attainable in general as retailers lack perfect knowledge on said preferences. In particular, an assortment strategy $\pi \in \mathcal{P}$ achieves in expectation a cumulative revenue given by:

$$J^\pi\left(F^{(\mathbb{N})}, T\right) := \mathbb{E}\Big\{ \sum_{t \leq T} r(S_t^\pi, F^t)\Big\},$$

where the expectation is taken over the series of (random) assortments $\left(S_t^\pi \ : \ t \leq T\right)$ offered by assortment strategy π. Because $F^{(\mathbb{N})}$ is not known by the retailer, we define the performance measure of an assortment strategy $\pi \in \mathcal{P}$ against the clairvoyant retailer, considering an adversarial realization of $F^{(\mathbb{N})}$. Specifically, we define the *adversarial regret* of an assortment strategy $\pi \in \mathcal{P}$ as:

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

The adversarial regret characterizes the worst-case opportunity cost incurred by the retailer when adopting an assortment strategy without full knowledge of preferences. A natural objective, therefore, is to construct a strategy that minimizes this regret. To formalize this idea, 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.

4 Dynamic assortment planning with evolving preferences

To shed light on the challenges of assortment planning under evolving preferences, in this section, we develop a structural characterization of preferences dynamics, with a particular focus on the magnitude of change. This characterization allows us to establish fundamental performance limits that no assortment strategy can surpass. Yet, we show that a simple restart-based strategy can match this limit, revealing not only how the retailer can achieve optimal performance, but also the inherent cost of operating in a dynamic and uncertain environment.

4.1 A macro-level view of customers' preferences

Rather than getting bogged down in fine-grained detail of customers' behavior, our focus is on capturing a broader, macro-level dynamic of evolving preferences. Accordingly, we introduce the concepts of *magnitude* and *velocity* to characterize the variability of the environment in which the retailer operates. Since we model preferences as probability distributions over random utility vectors, hereafter we use the Kullback–Leibler (KL) divergence as a measure of difference between customer's preferences (Thomas and Joy 2006).

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[2]. For $T \in \mathbb{N}$, we define the *magnitude of the environment* $\mathcal{M}(\mathcal{F}, T)$ as the highest cumulative variation along a sequence of customers' preferences among all possible such sequences. That is:

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

Thus, $\mathcal{M}(\mathcal{F}, T)$ measures the magnitude of potential changes that a retailer may face. Accordingly, a smaller magnitude guarantees that the retailer encounters only minor changes in preferences.

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:

$$p_i(S, F^t) := \frac{\nu_i^t}{\nu_0^t + \sum_{j \in S} \nu_j^t},$$

where $\big(\nu_i^t := \exp(\mu_i^t) \,:\, i \in [N] \cup \{0\}\big)$ are referred to as *attraction parameters* (Train 2009).

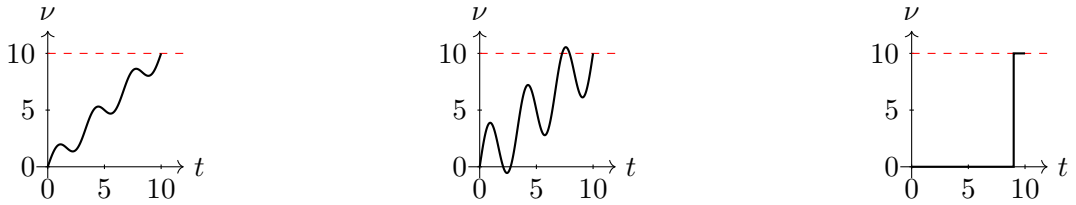
We define $\mathcal{F}_{\text{MNL}} \equiv \mathcal{F}_{\text{MNL}}(M_T)$, parameterized by $M_T > 0$, as the set of time-varying MNL models in which the attraction parameters ν_i^t switch between two regimes with equal probability at fixed intervals of length $\Delta = \lfloor T^{1/2}(8M_T)^{-1/2} \rceil$. Specifically, for $1 \le j \le \lceil T/\Delta \rceil - 1$, 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 as one can show 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 set $\mathcal{F}$ as the maximum difference in preferences between consecutive customers across all possible sequences of preferences. That is:

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

Thus, environments with a small velocity only admit gradual and slow changes in preferences. Conversely, environments with large velocity values allow for abrupt changes in preferences as well.



(a) Slow evolution of preferences. (b) Quick evolution of preferences. (c) Abrupt shift of preferences.

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, where a high value $\mathcal{V}(\mathcal{F}, T)$ results (Figures 1b and 1c) in a shift, while a lower value indicates (Figure 1a) a more gradual evolution.

Figure 1 illustrates this dynamic, showing how different velocities affect the possible changes in preferences. Note that the magnitude and velocity of the environment are always guaranteed to satisfy $\mathcal{M}(\mathcal{F}, T) \leq T \cdot \mathcal{V}(\mathcal{F}, T)$. Thus, for a fixed horizon T, one would expect lower velocities to be associated with a smaller magnitudes. In this study, we frame our discussion around the magnitude rather than the velocity, although similar insights could be derived for the latter.

4.2 A fundamental lower bound on the achievable performance

We establish a fundamental performance limit that applies to any assortment strategy. To derive this result, we construct a sequence of customers whose preferences are governed by the MNL choice model. Therefore, if $\mathcal{F}$ includes MNL preferences, then we obtain the following result.

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

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

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 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 "exploration," thereby pushing the regret upward.

Remark 1. The technical arguments used to obtain Theorem 1, can extend beyond the MNL choice model. Indeed, the same argument can be used to derive similar bounds for other random utility models commonly used in the literature. ■

If the magnitude of the environment 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. Therefore, our lower bound is larger than the classical $\mathcal{O}(\sqrt{T})$ regret achieved in settings with time-homogeneous preferences (Agrawal et al. 2019). Note that our lower bound differs from that of order $\mathcal{O}(T^{2/3}M_T^{1/3})$ obtained by Besbes et al. (2015) in non-stationary stochastic optimization (where M_T corresponds to a measure of the environment's variability). The key difference between their result and ours stems 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.

On the other hand, if the magnitude of the environment is in the order of $\mathcal{M}(\mathcal{F}, T) = \mathcal{O}(T^{\alpha})$ for some $\alpha \in (0, 1)$, then Theorem 1 establishes that the regret is bounded below by $\mathcal{O}\big(T^{\frac{3+\alpha}{4}}\big)$. As α approaches 1, the magnitude becomes linear in T and no policy can achieve sublinear adversarial regret. In other words, preferences might become so volatile that any attempt to adapt might be rendered ineffective in reducing the long-term worst-case opportunity cost.

4.3 A near-optimal restart-and-learn assortment strategy

We propose an assortment strategy inspired by Besbes et al. (2015), who develop a general framework to design policies in non-stationary stochastic optimization. While their approach is tailored to convex problems, we show that its core principles remain effective in our setting.

Algorithm 1 Restart-and-learn policy $\pi\,(\Delta, \mathcal{A})$

Input: A batch-size Δ and a policy $\mathcal{A}$ for the static setting

while $1 \leq j \leq \lceil T/\Delta \rceil$ **do**
  Run $\mathcal{A}$ on consumer $t = (j-1)\,\Delta + 1$ to $t = \min\big\{j\,\Delta, T\big\}$ (restart)
  $j = j + 1$

Recognizing the difficulty in precisely pinpointing shifts in consumer behavior, our approach ensures that the learning process is refreshed at regular intervals, allowing the retailer to update

its assortment according to the most recent estimates it has on customers' preferences. More precisely, the policy outlined in Algorithm 1 periodically restarts $\mathcal{A}$, an assortment strategy for time-homogeneous preferences (borrowed from the literature), at fixed intervals of Δ periods.

We assess the performance of Algorithm 1 in terms of the regret of $\mathcal{A}$ under the worst-case time-homogeneous preferences. Accordingly, we define the set of possible static preferences:

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

and denote the corresponding adversarial regret by $\mathcal{R}^{\mathcal{A}}(\mathcal{F}_S, T)$. Building on this worst-case opportunity cost, we next derive an upper bound on the regret of our assortment strategy in terms of both $\mathcal{R}^{\mathcal{A}}(\mathcal{F}_S, T)$ and the magnitude of the environment $\mathcal{M}(\mathcal{F}, T)$.

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 $\mathbf{w} \equiv \big(w_i : i \in \mathcal{N}\big)$. Then, for $T \geq 2$,*

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

To derive Theorem 2, we evaluate the regret of our assortment strategy in two steps. First, we introduce an intermediate benchmark—a "semi-oracle"—which fully knows customers' preferences in each Δ-length sub-segment but must offer a single assortment to all customers in that segment. Then, we derive the bound on the regret by measuring the revenue difference between our policy and the semi-oracle, and between the semi-oracle and the clairvoyant retailer.

Remark 2. The upper bound on regret relies on the performance of $\mathcal{A}$ under time-homogeneous preferences. Prior work in dynamic assortment planning, including Agrawal et al. (2019) for the MNL model and Li et al. (2025) for the MCC model, has established worst-case regret bounds for static preferences. These bounds, in turn, reflect the combinatorial complexity of the assortment planning problem. Thus, while Theorem 2 does not explicitly exhibit this combinatorial burden, it still implicitly depends on it through the regret incurred by $\mathcal{A}$ in the static setting. ■

The choice of Δ balances two competing forces. A smaller Δ allows for quicker adaptation to evolving preferences but increases reset frequency, limiting within-segment learning and leading to excessive exploration. Conversely, a larger Δ enables better preferences estimation but delays adaptation. Accordingly, Δ should be chosen small enough for timely response yet large enough to ensure meaningful learning within each segment.

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-and-learn algorithm achieves near-optimal performances, "almost" matching the lower bound on regret from Theorem 1. As both the horizon T and the magnitude $\mathcal{M}(\mathcal{F},T)$ increases, the volatility of the environment also increases. In response, reducing the size Δ ensures the algorithm adapts frequently enough to manage this increased volatility.

4.4 Operating with customers whose preferences evolve

The fundamental lower bound on achievable performance established in Section 4.2 illustrates the intrinsic challenge of operating in an environment with evolving preferences. To bring this theoretical observation to life, we present an example that varies the magnitude of the environment, revealing the direct impact on the performance of our assortment strategy.

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 set $\mathcal{F}_{\text{MNL}}(M_T)$ 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 strategy. ■

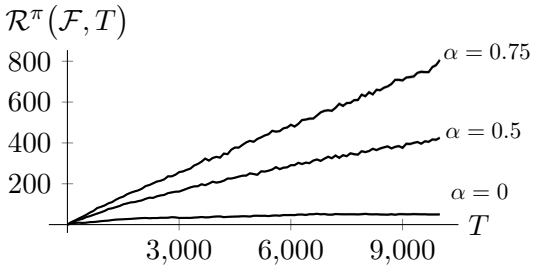


Figure 2: Regret of $\pi(\Delta,\mathcal{A})$ from Algorithm 1 applied to the sequence of settings from Example 2 with $M_T = T^{\alpha}$, for $\alpha \in \{0, 0.5, 0.75\}$, where the horizon ranges from $T = 1$ to $T = 10000$. The policy is described in Algorithm 1. We compute the average regret (in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

The insights from Figure 2 highlight a key challenge for retailers facing customers whose preferences evolve. As α increases, the magnitude of the environment M_T becomes larger, and therefore the oscillation between preferences from $\mathcal{F}_{\text{MNL}}(M_T)$ becomes more frequent (recall Example 1). This volatility limits the retailer's ability to stay up to date with the latest customers' preferences, thereby necessitating more frequent resets. Corollary 1 has a key economic consequence: in highly

variable environments, the inability to fully capture customers' preferences results in a significant opportunity cost, characterized by an increasing regret over time.

As described in Figure 2, if the magnitude of the environment increases, then the retailer faces greater challenges, resulting in higher regret. Yet, the restart-and-learn strategy is shown to be nearly optimal. However, it does not exploit any structural properties inherent in consumer behavior dynamics. This observation raises an important question: can the retailer develop an adaptive strategy that exploits any structural properties rather than relying solely on periodic resets?

5 Exploiting structural information about preferences' dynamics

In this section, we explore how structural information about potential changes can improve the retailer's performance. In particular, we consider settings in which the retailer expects a single abrupt change in preferences. Our analysis shows that detecting and responding effectively remains challenging in this situation. However, when external signals—such as market intelligence or internal analytics—offer additional insight into the magnitude, the retailer can act more proactively by attempting to detect the change through monitoring of purchasing data. Our analysis shows how access to structural information helps the retailer design more effective assortment strategies.

5.1 Information structure for abruptly changing preferences

In what follows, we assume that the retailer expects potential changes in preferences to occur abruptly. This belief may be informed by market analytics or expert judgment, especially in the context of significant disruptions, such as a pandemic, where both the timing and the impact of the change are uncertain. We model this situation by restricting our attention to a subset $\mathcal{F}_A \subseteq \mathcal{F}$ of preferences that are static except for a single unknown time $\tau \in \mathbb{N}$, which we define as follows:

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

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 "well separated" (Agrawal et al. 2019) 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 rather technical and prevents us from considering

settings in which learning post-change preferences becomes increasingly difficult as T grows.

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 any policy π via $\mathcal{R}^\pi(\mathcal{F}_A(F^1), T)$. In addition, we assume that any shift results in a change in the corresponding single-sale optimal assortment; that is, $S^*(F^1)$ and $S^*(F^\tau)$ must differ by at least one product.

5.2 Passively undetectable changes with few information on their magnitude

In this section, we consider environments characterized by an abrupt change, where preferences before and after the shift remain identical over the products included in the pre-change optimal assortment $S^*(F^1)$. As a result, a retailer continuing to offer this assortment would be unable to observe any change in purchasing behavior. Hence, we refer to such changes as *passively undetectable*. To capture this phenomenon, we introduce the following sub-class of preferences:

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

The condition in the definition of $\mathcal{F}_U$ ensures that customers' preferences cannot be statistically differentiated based solely on the information provided by the pre-change optimal assortment.

5.2.1 A fundamental lower bound on the achievable performance

We establish a fundamental lower bound on the performance of any assortment strategy when confronted with an abrupt and passively undetectable change in preferences. Since the magnitude of the environment (introduced in Section 4) serves somehow as the sole quantitative indicator available to the retailer, this lower bound is naturally expressed in terms of $\mathcal{M}(\mathcal{F}_U, T)$.

Proposition 1. *There exists some finite constant*[3] $C \equiv C(\gamma) > 0$*, such that, for* $T \geq 2$*:*

$$\widetilde{\mathcal{R}}\big(\mathcal{F}_U, T\big) \geq C\left(T^{1/2} \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1/2} - 1\right).$$

The regret bound in Proposition 1 reflects a fundamental trade-off when viewed through a game theoretic perspective of the interaction between the environment and the retailer. Because the retailer commits to a strategy in advance, a worst-case change can be delayed until the final period if exploration persists. Conversely, if exploration is limited at any time period, then the change may occur around that time. Proposition 1 shows that the lower bound decreases as the

magnitude increases. However, this observation does not imply that larger magnitude make the retailer's task any easier; rather, it suggests that the bound becomes somehow less informative.

Proposition 1 impose a constraint on the preferences that can be picked by the environment. By contrast, in Theorem 1, the environment adversarially selects preferences with very limited constraints (that define the set $\mathcal{F}$). In this regard, the environment is granted more flexibility in choosing preferences. From this perspective, abrupt changes represents a special case of the dynamic examined in Section 4, which in turn explains the difference in the achievable performance.

5.2.2 A near-optimal assortment strategy to face abruptly changing preferences

Settings with high magnitude may accommodate large or small changes. Accordingly, adapting to such changes may be challenging for the retailer. However, if the retailer knows that the magnitude is uniformly bounded above by a constant M, then the retailer benefits from knowing that preference changes might not be arbitrarily large. In this regime, using the restart-and-learn policy from Algorithm 1, with a carefully chosen segment length, matches the lower bound.

Proposition 2. *For $\mathcal{M}(\mathcal{F}_U, T) < M$, F^1 such that $F^{(\mathbb{N})} \in \mathcal{F}_U$, $\mathcal{A} \in \mathcal{P}$ and $\Delta \equiv \lceil T \cdot \mathcal{M}(\mathcal{F}_U, T) \cdot M^{-1} \rceil$, let $\pi \equiv \pi(\mathcal{A}, \Delta)$ be the policy defined in Algorithm 1. Then, for $T \geq 2$:*

$$\mathcal{R}^{\pi}(\mathcal{F}_U(F^1),\, T) \leq 4M \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1} \cdot \mathcal{R}^{\mathcal{A}}\left(\mathcal{F}_b(F^1), \Delta\right),$$

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

Our strategy partitions the horizon into $\mathcal{O}\big(M \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1}\big)$ segments, applying $\mathcal{A}$ repeatedly. If $\mathcal{A}$ achieves a regret of $\mathcal{O}(\sqrt{T \log T})$ under static preferences, as in Agrawal et al. (2019), then Proposition 2 implies that our strategy incurs a regret of order $\mathcal{O}\big(T^{1/2} \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1/2}\big)$, up to logarithmic terms. This regret "nearly" matches the lower bound in Proposition 1. Moreover, since our regret bound is strictly lower than that of Theorem 1, which is of order $\mathcal{O}(T^{3/4})$, it somehow quantifies the value of knowing that the change is abrupt and cannot be arbitrarily large.

Example 3. We consider a sequence of settings, in which the horizon T ranges from 1 to $50,000$, with an abrupt change at $\tau = T$. Preferences follow an MNL model with pre- and post-change attraction parameters ν^a and ν^b, respectively. For products $i \in \{1, 2, 3, 4\}$, we set $\nu_i^a = 0.25 + \zeta$, while all others have $\nu_i^a = 0.25$. After the change, products $i \in \{5, 6, 7, 8\}$ switch to $\nu_i^b = 0.25 + 1.1\zeta$, with all others remaining unchanged. We use $\zeta \in \{0.3, 0.35, 0.4\}$ and set $\nu_0^a = \nu_0^b = 1$ for the no-purchase option. Under this setup, the upper bound on the magnitude is $\mathcal{M}(\tilde{\mathcal{F}}_U, T) < (\zeta/5)^2$. The optimal assortment shifts from $S^*(\nu^a) = \{1, 2, 3, 4\}$ to $S^*(\nu^b) = \{5, 6, 7, 8\}$. These preferences

cannot be distinguished solely from purchases under $S^*(\nu^a)$, so that the induced sequence belongs to $\mathcal{F}_U(F^1)$. In this example, Algorithm 1 uses $\mathcal{A}$ by Agrawal et al. (2019) as a subroutine. ■

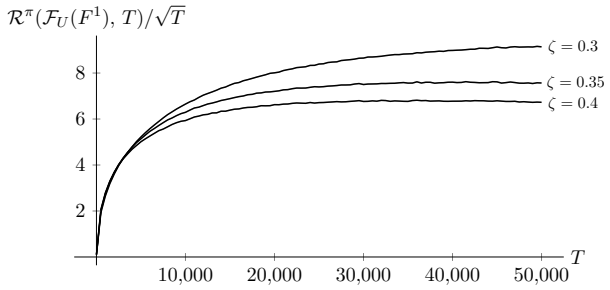


Figure 3: Regret of $\pi(\mathcal{A}, \Delta)$ from Algorithm 1, applied to the sequence of settings from Example 3 in which the horizon ranges from $T = 1$ to $T = 50,000$, with $\tau = T$. We compute the average regret (in black) and the 95% confidence interval for the mean (imperceptible), for $\zeta \in \{0.3, 0.35, 0.4\}$ over 500 instances.

As the magnitude decreases, regret increases due to the shrinking segment size Δ. This counter-intuitive observation occurs because smaller magnitude leads to minor, harder-to-detect changes, requiring extensive exploration for adaptation. Conversely, larger magnitude allows for both small and large changes, creating a trade-off: excessive exploration is necessary when there is no shift, while insufficient exploration makes the retailer "vulnerable" to larger changes, increasing the regret.

The retailer mitigates this trade-off by selecting a sub-segment size that is small enough to handle early large changes yet large enough to handle changes that are minor. Figure 3 highlights this trade-off and shows that regret increases as ζ decreases, aligning with Proposition 2. Indeed, lower values of ζ reduce the magnitude, which, in turn, decrease the segment size Δ. In such cases, our restart-and-learn strategy resets the learning process more frequently.

The discussion in this section highlights a key challenge in dynamic assortment planning with abrupt changes in preferences. The primary difficulty lies not only in large, sudden changes but also in smaller, more subtle ones. Overreacting to minor fluctuations may initially appear as excessive exploration. However, our findings demonstrate that ignoring subtle yet persistent changes can lead to long-term misalignment with customers' preferences. These observations underline the value of structural information, such as the abrupt nature of the change, in helping the retailer adapt more effectively. This raises a natural question: can additional knowledge about the nature or structure of the change further support the retailer in refining its assortment strategy?

5.3 Passively undetectable changes with information on their magnitude

We consider environments with abrupt and passively undetectable changes. Yet, we assume that the retailer expects the change to be neither arbitrarily small nor excessively large, but rather

to fall within a certain range. Accordingly, we introduce the subset of preferences defined by:

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

for some constants $\kappa \in (0,1)$ and $\phi > \kappa$, which provide a range for the magnitude $\mathcal{M}(\tilde{\mathcal{F}}_U, T)$. While the first condition above ensures that preferences are passively undetectable, the second one guarantees the existence of an assortment in which the preferences are sufficiently distinct. Together, these elements encapsulate the extra structural information available to the retailer, a property we refer to as the *separability* condition.

5.3.1 A fundamental lower bound on the achievable performance

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

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

This result follows the same reasoning as Proposition 1. Specifically, the argument makes a distinction between assortment strategies that explore sufficiently at all times and those that fail to do so, leading to the same fundamental trade-off.

5.3.2 An efficient active-learning assortment strategy

The separability condition provides additional information that enables the retailer to move away from restart-based approaches and instead adopt assortment strategies focused on detecting the change. Accordingly, we introduce the *active-monitoring-then-learn* policy (see Algorithm 2). The policy alternates between exploration and exploitation cycles of lengths $\Delta_e = \mathcal{O}(\log T)$ and $\Delta_o = \mathcal{O}(\sqrt{T})$, respectively. During each exploration cycle, the retailer offers assortments from $\mathcal{E}$, a subset of $\mathcal{S}$ designed to detect the change. If no change is detected via statistical testing, the pre-change optimal assortment is offered in the subsequent exploitation cycle. Otherwise, an assortment planning algorithm $\mathcal{A}$ is implemented for the remainder of the horizon to learn the new preferences.

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 4 bellow establishes an upper bound on the

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

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 $u = t+1, \ldots, t+\Delta_o$ (exploit pre-change optimal assortment)

Offer each assortment $S \in \mathcal{E}$ to Δ_e customers (explore)

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)

$t = t + \Delta_o + \left|\mathcal{E}\right|\Delta_e$

Run $\mathcal{A}$ on customers $t+1$ to T (post-change policy)

regret of our 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 the next section).

Proposition 4. *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 4 consists of three components, each capturing a distinct source of "inefficiency" in our assortment strategy. First, the $\mathcal{O}(\log T)$ term accounts for the detection delay induced by the statistical test used to identify preferences changes, reflecting the time required to gather sufficient evidence that a shift has occurred. Second, a $\mathcal{O}(\sqrt{T}\log T)$ term arises from continuously exploring assortments from $\mathcal{E}$, which is unnecessary in cases where the change happens near the end of the horizon. The last term corresponds to the regret incurred while learning the new optimal assortment after the change occurs.

Recall that Proposition 2 provides a regret bound in cases where the magnitude is uniformly bounded above. By contrast, Proposition 4 establishes a similar result under the additional assumption that the retailer knows that the change cannot be arbitrarily small. The value of this information manifests subtly within the $\mathcal{O}(\sqrt{T})$ term with $|\mathcal{E}|$. Indeed, note that such term is linear in the size of the set of test assortments $\mathcal{E}$, which captures the combinatorial structure of the problem through the number of test assortments. Consequently, it is important to construct $\mathcal{E}$ with minimal cardinality. In the next section, we construct this set by leveraging the separability

condition, thereby reducing the regret's dependence on the underlying combinatorial complexity.

5.3.3 Constructing a set of test assortments to detect changes

Constructing the set of test assortments $\mathcal{E}$ is critical for balancing exploration costs and detection accuracy 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.

(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[4] $p_i(S, \eta(\rho)) = \rho_i$ 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 parameters can be uniquely inferred from estimated purchasing probabilities, a property satisfied by the MNL (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*. As shown in Proposition 4, the regret bound scales with $|\mathcal{E}|$, growing as $\mathcal{O}\big(\binom{N}{K}\big)$ in the absence of structural assumptions. Assumptions (i) and (ii) ensure that any change in preferences affecting purchasing behavior is reflected in the model parameters. By following the partitioning approach, we can infer updated purchase probabilities, compare them to pre-change parameters, and detect changes in preferences. 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 $\tilde{\mathcal{F}}_U$. However, one could modify the statistical test for change detection in Algorithm 2 by:

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

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 $\tilde{\mathcal{F}}_U$.

When post-change preferences are known, it is possible to design a test assortment that balances detection performance and revenue exploitation (a direction briefly explored in Appendix A). The broader problem of selecting a set of test assortments, particularly those that better exploit available information about the magnitude of the change, remains an open question for future research.

Example 4. We consider a sequence of settings, in which the horizon T ranges from 1 to 100000, with a single abrupt change at $\tau = T$. Preferences follow an MNL choice model, with pre- and post-change attraction parameters ν^a and ν^b, respectively. Initially, products $i \in \{1, 2, 3, 4\}$ have $\nu_i^a = 0.6$, while others have $\nu_i^a = 0.1$. After the change, products $i \in \{5, 6, 7, 8\}$ switch to $\nu_i^b = 1$, with all others remaining unchanged. The no-purchase option has $\nu_0^a = \nu_0^b = 1$. The optimal assortment thus changes from $S^*(\nu^a) = \{1, 2, 3, 4\}$ to $S^*(\nu^b) = \{5, 6, 7, 8\}$. Moreover, $\kappa = 0.58$ is a valid lower bound for the maximum change magnitude in $\tilde{\mathcal{F}}_U$. The change is undetectable using only $S^*(\nu^a)$, ensuring the induced preferences $F^{(\mathbb{N})}$ belong to $\tilde{\mathcal{F}}_U$. ■

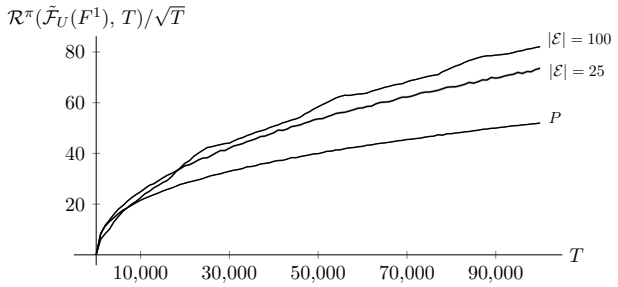


Figure 4: Regret of $\pi(\kappa, F^1, \mathcal{E}, \mathcal{A})$ from Algorithm 2 with $\kappa = 0.58$, applied to the sequence of settings from Example 4 in which horizon ranges from $T = 1$ to $T = 100,000$, where we set $\tau = T$. We compute the average regret (in black) and the 95% confidence interval for the mean (imperceptible) over 500 instances. Also, P indicates the partition approach whereas the value $|\mathcal{E}|$ indicates how many assortment were drawn randomly from $\mathcal{S}$.

We compare the partitioning approach P to a naive sampling one, where $|\mathcal{E}| \in \{25, 100\}$ assortments are selected randomly (as enumerating all possible assortment is computationally too expensive). The regret in Figure 4 evaluates our policy on the instance described in Example 4. When regret is scaled by $\sqrt{T}$, a logarithmic convergence pattern emerges, empirically validating Proposition 4. Moreover, these empirical results highlight that the size of $\mathcal{E}$ plays a key role in our assortment strategy performance: a smaller carefully chosen set leads to lower regret, emphasizing the importance of a strategic selection of the set of test assortments for the retailer.

5.4 Passively detectable changes in preferences

We now consider cases in which the retailer expects changes in customers' preferences to manifest themselves within the pre-change optimal assortment. For that purpose, we define the class of preferences with passively detectable changes as:

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

where we assume that the retailer possesses additional structural information regarding the change. Namely, the change cannot be arbitrarily small and $\varepsilon \in (0, 1)$ provides a lower bound on the

magnitude of the environment. Thus, preferences in $\mathcal{F}_D$ exhibits a large enough change toward products in the pre-change optimal assortment. Consequently, the retailer can detect such shifts "passively" by continuing to offer the pre-change optimal assortment and analyzing purchasing data.

5.4.1 A fundamental lower bound on the achievable performance

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 5. *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 construction argument in which we design an adversarial change point τ for any given assortment strategy. Specifically, for any policy, one can find τ such that the policy fails to offer the post-change optimal assortment within a time interval of length $\mathcal{O}(\log T)$ around τ, resulting in a regret increase of $\mathcal{O}(\log T)$. Remarkably, our bound, up to a constant factor, matches the classic lower bound by Lai and Robbins (1985) for well-separated multi-armed bandit problems. This finding highlights that when the retailer leverages structural detectability properties, identifying the change is as challenging as learning new well-separated customers' preferences.

5.4.2 An efficient passive-learning assortment policy

We introduce the *passive-monitoring-then-learn* policy, which leverages the detectability of the change in preferences (see Algorithm 3). This assortment strategy operates in cycles of length $\Delta = \mathcal{O}(\log T)$, during which the retailer monitors unexpected changes in purchasing frequencies within the pre-change optimal assortment. If no change is detected, then the pre-change optimal assortment is maintained; otherwise, the retailer switches to an algorithm to learn the new preferences. In contrast to Algorithm 2, there is no necessity in exploring alternative test assortments.

Next, we establish an upper bound on the regret of the passive monitoring assortment strategy. Proposition 6 shows that this bound, apart from the regret incurred when learning new preferences, closely matches the best achievable performance.

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

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 $u = t+1, \ldots, t+\Delta$ (exploit pre-change assortment)

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 6. *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 final term in the bound from Proposition 6 reflects the regret incurred by algorithm $\mathcal{A}$ in identifying the optimal assortment after the change. The initial term accounts for regret due to detection delays (or errors) from the statistical test. If the regret of $\mathcal{A}$ is $\mathcal{O}(\log T)$, as in the well-separated setting of Sauré and Zeevi (2013), then Proposition 6 shows that our policy also achieves $\mathcal{O}(\log T)$ regret. In essence, our strategy incurs significantly lower regret than the bound in Theorem 2. By leveraging the fact that the change is both abrupt and detectable (with a known magnitude range), the retailer can attain substantially lower opportunity costs compared to the general case in Section 4, where no such information is available.

Example 5. We consider a sequence of settings, in which the horizon T ranges from 1 to 10000, with an abrupt change at $\tau = 1$. Preferences follow an MNL models with pre- and post-change attraction parameters ν^a and ν^b, respectively. For products $i \in \{1,2,3,4\}$, we set $\nu_i^a = 0.25 + \zeta$ and for $i \in \{5,6,7,8\}$, $\nu_i^b = 0.25 + \zeta$, with $\zeta = 0.75$. Other products have $\nu_i^a = \nu_i^b = 0.25$, and the no-purchase option satisfies $\nu_0^a = \nu_0^b = 1$. Accordingly, the optimal assortment changes from $S^*(\nu^a) = \{1,2,3,4\}$ to $S^*(\nu^b) = \{5,6,7,8\}$. These preferences remain distinguishable under $S^*(\nu^a)$, and fixing $\varepsilon = 0.1$ ensures that the preferences are in $\mathcal{F}_D$. Moreover, applying the policy of Agrawal et al. (2019) for static preferences results in a regret of order $\mathcal{O}(\log T)$ for learning the new preferences, as our instances are well-separated (see Theorem 3 of their study). ■

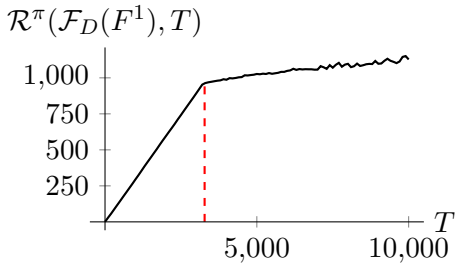


Figure 5: Regret of $\pi\left(\varepsilon, F^1, \mathcal{A}\right)$ from Algorithm 3, with $\varepsilon = 0.1$, applied to the sequence of settings from Example 5 in which the horizon ranges from $T = 1$ to $T = 10000$, where we set $\tau = 1$. We compute the average regret (in black), and 95% confidence interval for the mean (imperceptible) over 500 instances. Moreover, the vertical dashed line separates two regimes in the regret.

The regret in Figure 5 exhibits two distinct regimes. Initially, regret grows almost linearly as long as the horizon remains comparable to the cycle length Δ (recall Algorithm 3), the period used to collect consumer purchasing data. During this phase, the sample size is too small to draw reliable conclusions from the statistical test and to implement a new assortment. Once the horizon is large enough, the regret transitions to a logarithmic regime in T, supporting Proposition 6.

5.5 Leveraging structural information from abruptly changing preferences

The preceding sections illustrate how structural information about changes in customers' preferences can significantly improve the retailer's assortment strategy. Specifically, even limited information on the change (abrupt and uniformly bounded above) allows the retailer to reduce the regret's dependence on the number of customers from $\mathcal{O}(T^{3/4})$ in the generic case from Section 4 to $\mathcal{O}(T^{1/2})$. Beyond the benefits of well-timed restarts, our findings reveal that having structural insights about the abrupt change (such as a lower bound on the magnitude of the environment) can also help the retailer switch to a proactive monitoring based approach. By tracking deviations in preferences and adapting assortments only when necessary, the retailer avoids redundant exploration.

This advantage is particularly pronounced when the shift occurs within the pre-change optimal assortment. In such cases, collecting purchasing data from that assortment is sufficient to detect changes without disrupting operations. By leveraging this information on the detectability of the change, the retailer can introduce a new assortment as soon as a shift in preferences is detected, avoiding both excessive exploration and prolonged misalignment with new preferences. As a result, the incurred opportunity cost is only of order $\mathcal{O}(\log T)$ in addition to that of learning the new preferences. Notably, this regret scales far more favorably with the number of customers than the $\mathcal{O}(\sqrt{T})$ regret associated with passively undetectable environments.

6 Case study with data from a major Chilean retailer

In what follows, we illustrate the trade-offs discussed in the previous sections through a case study. To this end, we use clickstream data from a major Chilean retailer to simulate preferences reflecting two market scenarios. Specifically, we examine the evolution of preferences under (i) seasonal fashion diffusion and (ii) a sudden change induced by a pandemic shock. The first scenario is designed to corroborate that adapting to evolving preferences should yield better performance than applying an assortment strategy designed for static settings, something that might be obvious asymptotically but not necessarily in practice. The second scenario aims to illustrate the potential of incorporating structural information about the nature of the change to improve performances.

6.1 Implementation details

In our analysis, we use a clickstream dataset from a major Chilean retailer, comprising approximately 94,000 customer interactions. In each interaction, customers are presented with an assortment of $K = 4$ products drawn 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.

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 respective 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. Moreover, since the profit for each product is identical, the optimal assortments remain unchanged after scaling the attraction parameters[5]. All experiments were executed using Python 3.10 on a computing cluster equipped

with 32 Intel(R) Xeon(R) Gold 6126 CPUs (2.60 GHz) running Ubuntu 22.04.3; further details on our implementation can be found in Appendix B.

6.2 Adapting to changing preferences: is it worth it?

Description of the scenario. 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 (top row of Figure 6) ideal for winter conditions.

Figure 6: The top row shows three high boots (displayed in darker tones) from the initial optimal assortment. Over time, these are gradually replaced by three shorter shoes, as shown in the bottom row.

To construct our scenario, we postulate that as the season changes to summer, consumer tastes gradually evolve towards preferences for shorter, lighter shoes that offer improved comfort in warmer weather. We model this evolution as a gradual change from the initial preferences (as defined earlier) toward a new set of preferences characterized by attraction parameters $\hat{\nu}^T$.

To derive $\hat{\nu}^T$, we "swap" the attraction parameters for products from the top row of Figure 6 with those from the bottom row. Consequently, the optimal assortment at the last time period contains lighter shoes (those from the bottom row of Figure 6). The gradual "swap" in preferences throughout the horizon is illustrated in Figure 7, 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 $T = 2 \times 10^6$ customer visits to the retailer.

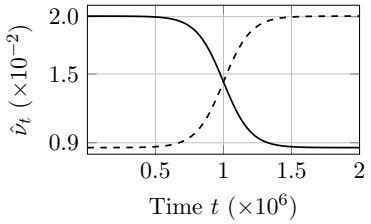


Figure 7: Attraction parameters evolution for Botas Argentinas (solid line) and Vince Camuto (dashed line). The evolution is governed by $s_t = (1 + \exp(-20\frac{t-100}{T-100} + 10))^{-1}$, with $T = 2 \times 10^6$; see Figure 6 for the product images.

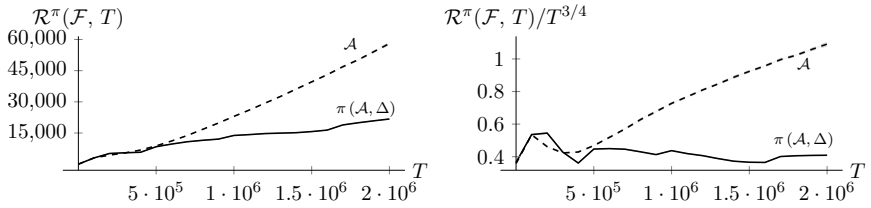


Figure 8: Regret incurred by two policies under customers' preferences that evolve from winter to summer preferences as a function of the horizon T. The *dashed* line corresponds to the policy $\mathcal{A}$ from Agrawal et al. (2019), whereas the *solid* line corresponds to the restart-and-learn policy $\pi(\mathcal{A}, \Delta)$ from Algorithm 1. We compute the average regret (in black), and 95% confidence interval for the mean (imperceptible) over 100 instances.

We compare two approaches for handling seasonal transitions: the assortment strategy from Agrawal et al. (2019) designed for time-homogeneous preferences (denoted by $\mathcal{A}$) and our restart-and-learn policy π from Algorithm 1. Policy π incorporates $\mathcal{A}$ as a subroutine and determines Δ based on Corollary 1. The regret for both approaches is presented in Figure 8. For the sake of comparison, both policies do not have access to the initial preferences $\hat{\nu}^1$. Moreover, we consider a sequence of settings in which the horizon T varies from $T = 1$ to $T = 2 \times 10^6$.

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 adaptive 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 periodic restarts) mitigate revenue erosion.

6.3 Exploiting structural information: does it really pay off?

Description of the scenario. We consider a setting in which customers' preferences change abruptly due to a pandemic, reminiscent of the COVID-19 crisis. In this scenario, we assume that the retailer has initially limited visibility into the impending surge in online shopping, unlike today's more advanced understanding of similar crises' effects on e-commerce (Oblander and McCarthy 2023). We calibrate the pre-change preferences using the data from the 30–39 age range, represented by the attraction parameters $\hat{\nu}^1$ in Figure 9. When the pandemic strikes, a broader cross-section of consumers transitions from in-store to online shopping (from one day to the next), leading to new attraction parameters $\hat{\nu}^\tau$, calibrated using the entire data.

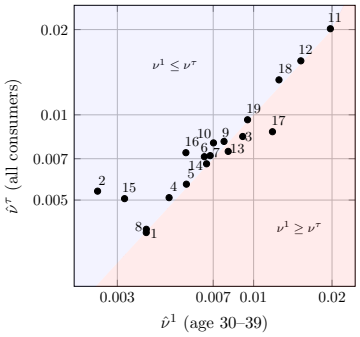
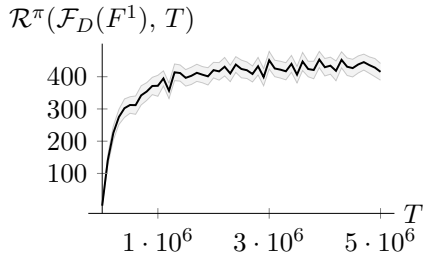
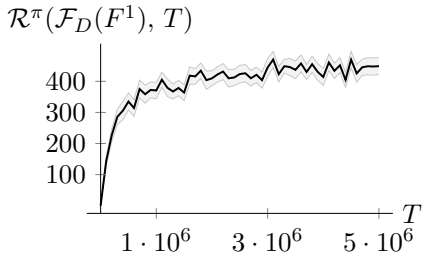


Figure 9: Estimated attraction parameters $\hat{\nu}^1$ (age 30–39) against $\hat{\nu}^\tau$ (all consumers).

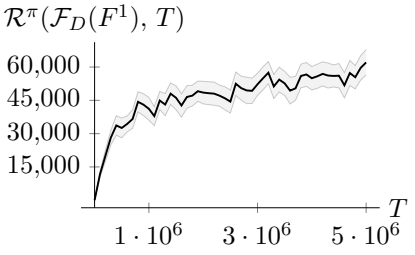
The optimal assortment before the change occurs is $S^*(\hat{\nu}^1) = \{11, 12, 17, 18\}$, which then shifts to $S^*(\hat{\nu}^\tau) = \{11, 12, 18, 19\}$. Additionally, we consider a sequence of settings in which the horizon ranges from $T = 1$ to $T = 5 \times 10^6$, with 100 independent replications 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 $\nu^t = \hat{\nu}^\tau$.



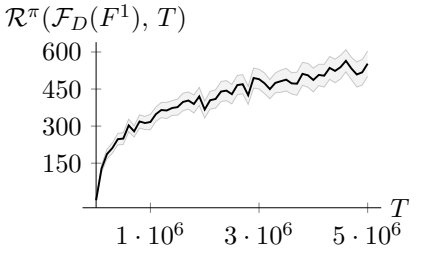
(a) Regret of the Algorithm 3.



(b) Regret of Algorithm 6 (in Appendix A).



(c) Regret of Algorithm 2.



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

Figure 10: 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 10a and Figure 10c show the regret of the passive and active monitoring-then-learn policies, respectively, when the post-change preferences are unknown. Similarly, Figure 10b and Figure 10d present the regret of the passive and active monitoring-then-learn policies, respectively, when the post-change preferences are known.

We compare four algorithms designed for abrupt changes in preferences: Algorithms 2 and 3 address cases where the retailer has very limited prior knowledge on the post-change preferences. In contrast, Algorithms 4 and 6, detailed further in Appendix A, are based on the assumption that the retailer has full knowledge of these new preferences. Moreover, this scenario is such that the change is passively detectable as it affects the attraction parameters of the products within the

pre-change optimal assortment. As a result, Algorithms 3 and 6 are both applicable in this setting.

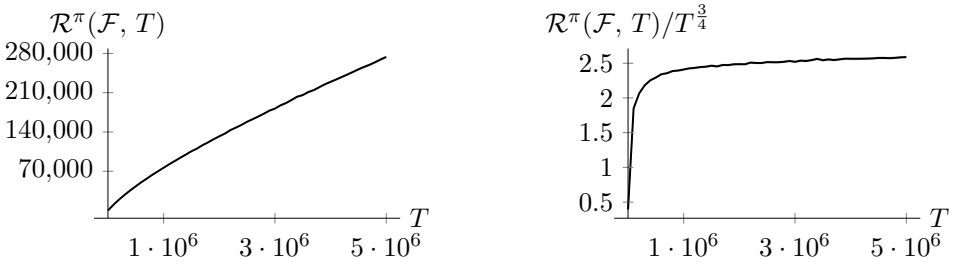
To facilitate a fair comparison among these algorithms, we deliberately exclude the regret related to the learning phase of the preferences once the change is detected. This methodological choice allows us to isolate the opportunity cost of detecting preferences shifts and examine whether strategically leveraging the structure of the change offers meaningful advantages for the retailer. Figure 10 illustrates the regret associated with each of these four algorithms.

Discussion. Our findings reveal that the passive assortment strategy presented in Algorithm 3, which does not presume exact knowledge of the post-change preferences, yields a performance (Figure 10a) nearly equivalent to the passive strategy assuming complete knowledge from Algorithm 6 (Figure 10b). Comparing the two, we observe that the retailer incurs only a minor performance loss from not knowing the post-change preferences (provided that the change can be detected passively). Note that the effectiveness of Algorithm 3 depends on the separability parameter, which implicitly reflects some prior knowledge about the magnitude of the environment.

In contrast, when the retailer is unaware that the change occurs within the pre-change optimal assortment and adopts an active exploration strategy as described in Algorithm 2, a large increase in regret is observed (Figure 10c). Under these conditions, the necessity of engaging in active exploration results in greater regret compared to the passive strategies, which can be explained by the frequency of exploration (whose batches size are of order of $\mathcal{O}(\sqrt{T})$ in the active strategy). Accordingly, when a change occurs, the policy may first have to end the exploitation batch, and then start an exploration one before detecting the change. This in turn drives the regret upward.

Moreover, Algorithm 4, which is designed for known post-change preferences, exhibits an improved performance (Figure 10d) compared to its counterpart, Algorithm 2 (Figure 10c). This gain has two sources. First, Algorithm 2 incurs higher regret due to broader exploration as it must offer all assortments from the set of test assortments, while Algorithm 4 can focus on a single one, reducing both exploration and regret. Second, their statistical tests differ: Algorithm 2 relies on estimated purchase probability gaps and requires a minimum sample size, whereas Algorithm 4 uses a more "efficient" likelihood ratio test. These distinctions are captured in Proposition 6, where the regret bound only applies for sufficiently large T.

Next, we examine a situation in which the retailer does not anticipate an abrupt change in preferences and adopts the restart-and-learn policy from Algorithm 1. This policy incorporates the assortment strategy $\mathcal{A}$ from Agrawal et al. (2019), with Δ chosen as in Corollary 1. As depicted in Figure 11a, the regret incurred by this policy is higher compared to the strategies specifically de-



(a) Regret of Algorithm 1.

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

Figure 11: Regret (and its scaled version) for the restart-and-learn policy $\pi(\mathcal{A}, \Delta)$ described in Algorithm 1. We compute the average regret (in black), and 95% confidence interval for the mean (imperceptible) over 100 instances.

signed for abrupt changes in preferences, shown in Figure 10. However, caution should be exercised in directly comparing regret values, since the opportunity cost associated with learning post-change preferences has been omitted from the analyses of abrupt-change-specific algorithms. Yet, the regret of Algorithm 1 grows at a rate of $\mathcal{O}(T^{3/4})$, a trend we observe in Figure 11b. Collectively, these observations highlight the strategic benefit of leveraging structural insights on preferences.

To summarize, our findings provide a significant insight for retailers anticipating an abrupt change in preferences. Focusing on relatively simple detection procedures, as opposed to restart-and-learn policy, improves performance by enabling the retailer to rapidly move away from obsolete assortments. This advantage is particularly pronounced in scenarios in which disruptive events shift customers' preferences away from the products offered within the pre-change optimal assortment. Overall, our analysis demonstrates that **retailers can significantly reduce the regret by proactively implementing change detection methods and capitalizing on structural information regarding the change in preferences**.

7 Conclusion

Saving retailers from the boiling market. Customers with evolving preferences pose a challenge to any retailer, particularly when considering the opportunity costs involved. By being confronted to these changing preferences, retailers risk a fate akin to the boiling frog—gradually missing vital market changes until it becomes too late. Our analysis reveals that operating in a dynamic market carries an inherent opportunity cost: a premium for delayed adaptation or inaction. Nonetheless, we offer insights on designing an assortment strategy to operate in this dynamic market. By periodically restarting their learning process (essentially reapplying a policy tailored for time-homogeneous preferences) retailers can avoid reliance on outdated preferences and, in doing so, sidestep the peril of becoming the unwitting frog in a boiling market.

If retailers gain external insights into the dynamics of customers' preferences—including the velocity, magnitude, and detectability of change—then they can leverage this information to refine their assortment strategies. In our study, we focus on scenarios in which an abrupt shift in preferences is anticipated by the retailer. When the magnitude of change is unknown in advance, retailers can preserve their adaptability by regularly resetting their learning process. Conversely, if market intelligence or expert insights indicate a sudden and large change in preferences, then retailers can preemptively embed change-detection mechanisms into their assortment strategies. This proactive approach mitigates the opportunity cost especially when such changes are passively detectable.

Managerial implications. Our findings highlight the importance of moving beyond assortment planning designed for time-homogeneous customers' preferences and adopting strategies that adapt to dynamic environments. In such markets, external information plays a crucial role in helping retailers refine their assortment strategies. Equally critical is achieving the appropriate balance between exploration and exploitation. Retailers should allocate resources for exploratory assortments designed to proactively identify emerging preferences without jeopardizing revenue streams. Striking this balance allows retailers to avoid missed opportunities at limited costs. Ultimately, retailers must discard the assumption that any assortment, once set, can remain indefinitely effective. Success in contemporary retail markets demands continuous vigilance and a proactive approach to refining assortments that resonate with consumers' ever-changing preferences.

Future research. Looking ahead, a key challenge lies in developing test assortments that can more effectively detect abrupt changes in preferences. While our model of preferences is quite general, introducing more structure, as for example adopting a specific choice model would allow to leverage this structure to design test assortments to mitigate the opportunity costs of delayed adaptation. Further improvements in assortment planning can be achieved by integrating contextual information—such as macroeconomic trends, demographic changes, or social media sentiments—to enhance strategy effectiveness. For instance, insights from one market might inform assortment decisions in another (Elberse and Eliashberg 2003). Additionally, incorporating realistic consumer behaviors, including seasonality effects (Caro et al. 2014) and brand loyalty (J. N. Sheth 1967), would help bridge the gap between theoretical research and practical retail applications.

Notes

[1]For accuracy, a frog with a brain starts being agitated when the temperature reaches 25°C (Lewes 1873).

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

[3]Most of our performance bounds exhibit a dependence on a constant C. This constant should not be interpreted as having the same value across all results. For the specific form of the constant and its dependence on the parameter settings, please refer to the proofs of the corresponding results.

[4]By abuse of notation, $p(\cdot, \theta)$ denotes the purchasing probability of the distribution parameterized by some θ.

[5]This observation follows from that $x \to \frac{x}{1+x}$ is an increasing function over $(0, +\infty)$.

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.

Bertsimas, D. and Mišić, V. V. (2019). "Exact first-choice product line optimization". In: *Operations Research* 67.3, pp. 651–670.

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.

Caro, F., Martínez-de-Albéniz, V., and Rusmevichientong, P. (2014). "The assortment packing problem: Multiperiod assortment planning for short-lived products". In: *Management Science* 60.11, pp. 2701–2721.

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

Chen, X., Ma, W., Simchi-Levi, D., and Xin, L. (2024). "Assortment planning for recommendations at checkout under inventory constraints". In: *Mathematics of Operations Research* 49.1, pp. 297–325.

Chintagunta, P. K., Bonfrer, A., and Song, I. (2002). "Investigating the effects of store-brand introduction on retailer demand and pricing behavior". In: *Management Science* 48.10, pp. 1242–1267.

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.

Elberse, A. and Eliashberg, J. (2003). "Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures". In: *Marketing Science* 22.3, pp. 329–354.

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.

Garivier, A. and Moulines, E. (2011). "On upper-confidence bound policies for switching bandit problems". In: *International conference on algorithmic learning theory*. Springer, pp. 174–188.

Golrezaei, N., Nazerzadeh, H., and Rusmevichientong, P. (2014). "Real-time optimization of personalized assortments". In: *Management Science* 60.6, pp. 1532–1551.

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.

Hoch, S. J. and Loewenstein, G. F. (1991). "Time-inconsistent preferences and consumer self-control". In: *Journal of Consumer Research* 17.4, pp. 492–507.

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.

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

Lai, T. L. (1998). "Information bounds and quick detection of parameter changes in stochastic systems". In: *IEEE Transactions on Information Theory* 44.7, pp. 2917–2929.

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

Lattin, J. M. (1987). "A model of balanced choice behavior". In: *Marketing Science* 6.1, pp. 48–65.

Lewes, G. (1873). "Sensation in the Spinal Cord". In: *Nature* 9, pp. 83–84.

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.

Oblander, S. and McCarthy, D. M. (2023). "Frontiers: estimating the long-term impact of major events on consumption patterns: evidence from COVID-19". In: *Marketing Science* 42.5, pp. 839–852.

Pollak, M. (1985). "Optimal detection of a change in distribution". In: *The Annals of Statistics*, pp. 206–227.

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.

Shiryaev, A. N. (1963). "On optimum methods in quickest detection problems". In: *Theory of Probability & Its Applications* 8.1, pp. 22–46.

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.

Tartakovsky, A. G. and Veeravalli, V. V. (2005). "General asymptotic Bayesian theory of quickest change detection". In: *Theory of Probability & Its Applications* 49.3, pp. 458–497.

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

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.

Tucker, C. (2008). "Identifying formal and informal influence in technology adoption with network externalities". In: *Management Science* 54.12, pp. 2024–2038.

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.

Wald, A. and Wolfowitz, J. (1948). "Optimum character of the sequential probability ratio test". In: *The Annals of Mathematical Statistics*, pp. 326–339.

Wood, W. and Neal, D. T. (2009). "The habitual consumer". In: *Journal of Consumer Psychology* 19.4, pp. 579–592.

Zhang, J., Ma, W., and Topaloglu, H. (2024). "Leveraging the degree of Dynamic Substitution in Assortment and Inventory Planning". In: *Operations Research*. To appear.

Zhou, H., Wang, L., Varshney, L., and Lim, E.-P. (2020). "A near-optimal change-detection based algorithm for piecewise-stationary combinatorial semi-bandits". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 04, pp. 6933–6940.

Appendix A The value of known post-change preferences

In this section, we discuss a setup similar to the one from Section 5 with abruptly changing preferences. However, the retailer is now assumed to know the post-change preferences. 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 assortment strategies. Furthermore, this setting serves as a 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 environment is bounded above uniformly. This assumption is equivalent to assuming that the probability of purchase for any product cannot be made arbitrarily small by the environment.

A.1 Passively undetectable changes

We consider passively undetectable changes as first discussed in Section 5.2. Accordingly, 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.1.1 A fundamental lower bound on the achievable performance

We establish a lower bound on the regret that any admissible policy 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 explore 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 regret.

Proposition 7. *There exists some finite constant $C \equiv C(\gamma, \vartheta) > 0$, such that, for $T \geq 2$:*

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

The lower bound in Proposition 7 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 3, where the post-change preferences are unknown, and the difference between the two results lies in the constant in front of the term $\sqrt{T}$. Hence, the regret incurred by any policy, whether due to learning static preferences (Agrawal et al. 2019) or detecting a change, remains of the same order in the worst-case.

A.1.2 A near-optimal assortment strategy

We specialize the active-monitoring-then-learn policy from Algorithm 2 to the setting in which the post-change preferences F^τ are known, as described in Algorithm 4. This specialization manifests in two key aspects. First, the algorithm leverages F^τ by performing a log-likelihood ratio test 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-then-optimize policy $\pi\left(D, F^1, F^\tau, S\right)$

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 $u = t+1, \ldots, t+\Delta_o$ (exploit pre-change assortment)
  Offer assortment S to Δ_e customers
  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 8, 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 8. *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 8 aligns with the lower bound in Proposition 7, 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. An inappropriate choice may lead to a high value of C in Proposition 8 and therefore to higher regret in the worst case. We briefly address the selection of such assortment in Section A.1.3.

Example 6. We consider a sequence of settings in which the horizon T ranges from 1 to 10000 with $\tau = T$. Customers' preferences follow an MNL model with $N = 10$ products and $K = 4$, where we set $w_i = 1$ for all products. The experiment is repeated over 500 randomized instances. The attraction parameters are set as follows: $\nu_0^a = \nu_0^b = 1$, and for $i \in \{1, 2, 9, 10\}$, we set $\nu_i^a = 0.25 + \zeta$, and $\nu_i^a = 0.25$ otherwise. Similarly, $\nu^b = 0.25 + 2\zeta$, for $i \in \{3, 4, 5, 6\}$, and $\nu_i^b = 0.25$ otherwise, with $\zeta = 2.75$. The Type I and Type II error probabilities are both controlled using $\alpha_{\mathrm{I}} = \alpha_{\mathrm{II}} = 0.01$. The two preferences differ only for products 3 to 6, with the pre-change and post-change optimal assortments given by $S^*(F^1) = \{1, 2, 9, 10\}$ and $S^*(F^\tau) = \{3, 4, 5, 6\}$, respectively. Thus, the preferences are guaranteed to belong to the subset $\mathcal{F}_U$. ■

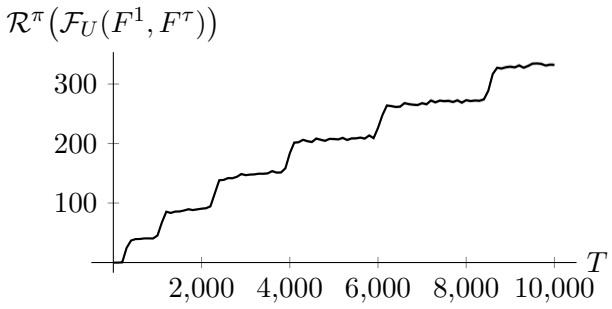


Figure 12: Regret of the active-monitoring-then-optimize policy $\pi\left(D, S, F^1, F^\tau\right)$ applied to the sequence of settings from Example 6 in which the horizon T ranges from $T = 1$ to $T = 10000$ with $\tau = T$. We compute the average regret (in black), and the 95% confidence interval for the mean (imperceptible) over 500 instances. The policy controls the Type I and II errors at $(\alpha_{\mathrm{I}}, \alpha_{\mathrm{II}}) = (0.01, 0.01)$. Also, $S := S^*(F^\tau)$ is used as the test assortment.

The upper bound on the regret of Algorithm 4 from Proposition 8 exhibits the same dependence on T as the bound derived for the case in which the post-change preferences are unknown (see

Proposition 4). Specifically, the regret is of order $\mathcal{O}(\sqrt{T}\log T)$, a trend that is empirically supported by the results presented in Figure 12. The step-wise pattern observed in the figure arises from the structure of the policy, which partitions the selling horizon into exploratory sub-segments. As T increases, the number of these segments grows, thereby contributing incrementally to the regret.

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.1.3 Finding a test assortment

In this section, we discuss approaches for selecting a "good" test assortment S to be used in Algorithm 4. Specifically, a test assortment should balance the expected revenue maximization and the ability of the retailer to detect the change based on the information collected from that assortment. Solving this trade-off at optimality in its full generality is challenging and beyond the scope of this paper. Instead, we propose evaluating candidate assortments based on two key criteria: (i) *separability*, i.e., the retailer may prefer assortments that maximize the statistical separation between the pre- and post-change preferences, thereby aiming for faster change detection; (ii) *short-term revenue*, i.e., the assortment may be selected to maximize a single-period expected revenue, thus aiming to reduce regret in the event that no change occurs.

Separability. We propose an approach to find an assortment that maximizes the difference between the pre- and post-change preferences. To proceed, we introduce a subset of feasible assortments $\Lambda \subseteq \mathcal{S}$. We then define a parametric optimization problem to find an assortment $S \in \Lambda$ that maximizes the worst-case separability between the pre- and post-change preferences. Formally:

$$z_{\text{SEP}}(\Lambda) := \max_{S\in\Lambda} \min_{i\neq j\in\{1,\tau\}} \mathcal{K}(F^i\,,\,F^j\,;\,S).$$

Note that the KL divergence is not symmetric (Thomas and Joy 2006); thus, worst-case separability refers to the smallest divergence between the pre- and post-change preferences, in either direction.

Short-term revenue. The second proposed approach aims to identify an assortment that can distinguish between pre- and post-change preferences, but with a focus on minimizing a single-period regret. To proceed, we introduce a subset of feasible assortments $\Lambda \subseteq \mathcal{S}$. We then define a parametric optimization problem that selects an assortment $S \in \Lambda$, which minimizes the worst-case regret (with respect to either F^1 or F^τ), while ensuring that the two preferences remain

distinguishable under the chosen assortment. Specifically:

$$z_{\text{REV}}(\Lambda) := \min_{S\in\Lambda}\Big\{ \max_{F\in\{F^1,F^\tau\}}\{r(S^*(F),F) - r(S,F)\} \,:\, \|p(S,F^1) - p(S,F^\tau)\|_\infty > 0\Big\}.$$

In words, the formulation above selects an assortment that minimizes the worst-case regret for a single period while still being able to differentiate the pre- and post-change preferences based on the information collected by offering that assortment.

Enumerating all assortments to solve either $z_{\text{SEP}}(\mathcal{S})$ or $z_{\text{REV}}(\mathcal{S})$ would return an optimal assortment with the desired properties (optimizing either separability or short-term revenue). However, such an exhaustive approach can be impractical due to the combinatorial size of $\mathcal{S}$. To address this issue, we limit our attention to an heuristic approach in order to derive a feasible solution to these programs. In particular, we propose a local-search type of approach described in Algorithm 5.

Algorithm 5 Find test assortment $\mathcal{T}(S^0, K, z)$

Input: An assortment S^0, an assortment size K and parametric program $z : 2^{\mathcal{S}} \to \mathbb{R}_{\geq 0}$
for $k \in [K]$ **do**
  $\Lambda = N_k(S^0)$
  Compute $z(\Lambda)$ (solve parametric program)
  if $z(\Lambda) > 0$ **then** (test assortment detected)
    Find the corresponding optimal assortment S^* (informative assortment)
    Stop and return S^*
Return: S^*

Algorithm 5 starts from an initial assortment $S^0 \in \mathcal{S}$ and explores its k-flip neighborhood for $k \in [K]$. The k-flip neighborhood is defined as $N_k(S^0) := \{S \in \mathcal{S} \,:\, \|S - S^0\|_1 = k\}$, representing all assortments that differ from S^0 by exactly k products. These assortments are obtained by replacing items from S^0 with items not currently in the assortment. If no suitable assortment is found in the current neighborhood, then k is incremented and the search continues until an "informative" assortment is identified. As k increases, the search space expands and eventually satisfies $N_K(S^0) = \mathcal{S}$.

Lemma 1. *For $S^0 \in \mathcal{S}$, $K > 0$, $z \in \{z_{SEP}, z_{REV}\}$, let $\mathcal{T} \equiv \mathcal{T}(S^0, K, z)$ denote the procedure as specified in Algorithm 5. Then, the following properties hold: (i) $\mathcal{T}$ returns a feasible assortment $S^* \in \mathcal{S}$ such that $z(\{S^*\}) > 0$; (ii) $\mathcal{T}$ terminates in a finite number of steps, with a worst-case computational complexity of order $\mathcal{O}(N^K)$.*

Lemma 1 establishes theoretical guarantees for the procedure described in Algorithm 5. Specifically, the algorithm is guaranteed to produce an assortment that satisfies the requirements of the

assortment strategy described by Algorithm 4. In the next example, we illustrate the numerical performances for the two approaches.

Example 7. We consider a retailer offering $N = 10$ products, each with equal profit $w_i = 1$, and we set $T = 10000$. Preferences follow an MNL model, with an abrupt change occurring at $\tau \in \{0, 2500, 5000, 7500, 10000\}$. Each experiment is randomized over 500 instances. We fix $\nu_0^1 = \nu_0^\tau = 1$. For products $i \in \mathcal{N}$, we set $\nu_i^1 \sim \text{Uniform}(0.1, 1)$. Then, we use $\nu_i^\tau \sim \text{Uniform}(1, 2)$, for all products except for those in $S^*(F^1)$, where we set $\nu_i^\tau = \nu_i^1$. The Type I and II errors are controlled at level $\alpha_{\text{I}} = \alpha_{\text{II}} = 0.1$. We compute the regret and the delay of detection (i.e., time at which the change is detected minus the time of change τ) for Algorithm 4, using the test assortment selected by Algorithm 5, where we use $S^0 = S^*(F^1)$ as an initial assortment and $z \in \{z_{\text{SEP}}, z_{\text{REV}}\}$. The results for the regret are shown in Table 2 and those for the delay in Table 3. ■

z \ τ	0	2500	5000	7500	10000
z_{SEP}	**370** (±17)	**201** (±15)	**211** (±13)	**183** (±9)	34 (±4)
z_{REV}	641 (±28)	397 (±25)	304 (±17)	200 (±8)	**12** (±4)

Table 2: Regret achieved by Algorithm 4 when applied to the setting of Example 7. The test assortment used in the assortment strategy is obtained using $\mathcal{T}(S^0, K, z)$ for $S^0 = S^*(\nu^1)$ and $z \in \{z_{\text{SEP}}, z_{\text{REV}}\}$ as described in Algorithm 5. The table reports the mean regret across 500 instances, with 95% confidence interval for the mean in parentheses for various time change $\tau \in \{0, 2500, 5000, 7500, 10000\}$. The lowest regret is shown in bold.

Table 2 indicates that using z_{SEP} as a subroutine for $\mathcal{T}$ leads to lower regret over the horizon compared to using z_{REV}. These experiments suggest that optimizing for separability (i.e., rapid change detection) can lead to improved long-term performance. As τ increases, however, the difference in regret between the two approaches diminishes. Notably, in a special case in which no change occurs ($\tau = T$), optimizing short-term revenue via z_{REV} results in lower regret than optimizing separability. This outcome is intuitive, as the strategy ends up exploring "for nothing" when no change occurs. As a result, choosing an assortment that minimizes single-period regret (such as the one selected by z_{REV}) becomes the preferable option. On the other hand, Table 3 shows that selecting the test assortment using z_{SEP} prioritizes separability and improves detection performance. Indeed, we observe shorter detection delays under z_{SEP} compared to z_{REV}.

We do not claim that one approach—optimizing revenue via z_{REV} or optimizing separability via z_{SEP}—is universally superior. Rather, our admittedly limited experiments with Example 7 illustrate that each strategy can perform well in some environments. Moreover, we did not observe this pattern across all instances. In some cases, optimizing one objective unexpectedly led to better

z \ τ	0	2500	5000	7500	10000
z_{SEP}	**3682** ($\pm$199)	**2080** ($\pm$174)	**2022** ($\pm$129)	**1553** ($\pm$72)	0 ($\pm$0)
z_{REV}	6269 ($\pm$267)	4042 ($\pm$252)	3001 ($\pm$151)	1946 ($\pm$68)	0 ($\pm$0)

Table 3: Delay achieved by Algorithm 4 when applied to the setting of Example 7. The test assortment used in the assortment strategy is obtained using $\mathcal{T}(S^0, K, z)$ for $S^0 = S^*(\nu^1)$ and $z \in \{z_{\text{SEP}}, z_{\text{REV}}\}$ as described in Algorithm 5. The table reports the mean regret across 500 instances, with 95% confidence interval for the mean in parentheses for various time change $\tau \in \{0, 2500, 5000, 7500, 10000\}$. The lowest regret is shown in bold.

outcomes for the other. Investigating further the question of finding a "good" test assortment presents an interesting direction for future research that we aim to explore in the future.

A.2 Passively detectable changes in preferences

In this section, we consider a setting in which the retailer knows that changes can be detected passively, i.e., $F^{(\mathbb{N})} \in \mathcal{F}_D$. 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 5.4, we assume that the retailer knows the post-change preferences F^τ.

A.2.1 A fundamental lower bound 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 9. *There exists constants $C \equiv C(\gamma, \vartheta) > 0$ and $t \equiv t(\gamma, \vartheta) > 0$, such that, for $T \geq t$:*

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

Proposition 9 establishes a regret lower bound in the setting where the post-change preferences are known. This result parallels that of Besbes and Zeevi (2011), who study dynamic pricing under abrupt demand shifts with known post-change demand. Since having access to additional information cannot degrade performance, the lower bound from Proposition 9 remains valid in the unknown-preferences setting discussed in Proposition 5.

A.2.2 An efficient passive-learning assortment policy

Next, we present a variation of the passive-monitoring-then-learn policy 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.

Algorithm 6 Passive-monitoring-then-optimize policy $\pi\left(D, F^1, F^\tau\right)$

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 $u = t+1, \ldots, t+\Delta$ (exploit pre-change assortment)

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 6, 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 10 which provides an upper bound for the regret of Algorithm 6.

Proposition 10. *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 6. 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 9 up to a constant. 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 6 when the post-change preferences are unknown is in the same order as the one from Proposition 10. In the next example, we provide a numerical illustration of the performance of our policy from Algorithm 6.

Example 8. We consider a sequence of settings in which the time horizon T ranges from 1 to 10000, with $\tau = 1$. Customer preferences follow an MNL model with $N = 10$ products and $K = 4$, where $w_i = 1$ for all products. Results are averaged over 500 randomized instances. The attraction parameters are set as follows: $\nu_0^a = \nu_0^b = 1$, and for $i \in \{1, 2, 9, 10\}$, we set $\nu_i^a = 0.25 + \zeta$, and $\nu_i^a = 0.25$ otherwise. Similarly, $\nu_i^b = 0.25 + \zeta$ for $i \in \{3, 4, 5, 6\}$, and $\nu_i^b = 0.25$ otherwise, with $\zeta = 2.75$. The Type I and II errors are controlled using $\alpha_{\mathrm{I}} = \alpha_{\mathrm{II}} = 0.01$. The optimal assortments before and after the change are $S^*(F^1) = \{1, 2, 9, 10\}$ and $S^*(F^\tau) = \{3, 4, 5, 6\}$, respectively. Thus, the resulting preferences are in $\mathcal{F}_D$. ■

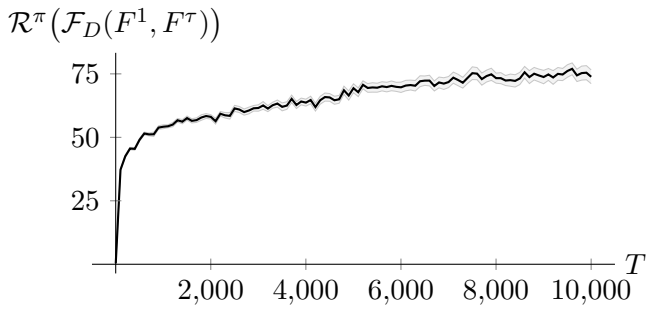


Figure 13: Regret of the passive-monitoring-then-optimize policy $\pi\left(C, F^1, F^\tau\right)$ applied to the sequence of settings from Example 8 in which the horizon T ranges from $T = 1$ to $T = 10000$, where we set $\tau = 1$. We compute the average regret (in black), and the 95% confidence interval for the mean (in shading) over 500 instances. The policy controls the Type I and II errors at $(\alpha_{\mathrm{I}}, \alpha_{\mathrm{II}}) = (0.01, 0.01)$.

The regret of the passive-monitoring-then-optimize policy, as shown in Figure 13, increases logarithmically with T. This empirical observation is consistent with Proposition 10, which establishes a logarithmic regret bound when the post-change preferences are known. Together, Figure 13 and Proposition 10 show that having access to the post-change preferences influences only the leading constant of order $\log T$ in the regret, without improving the rate of growth in T.

Appendix B 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 assortment strategy, we fix the value taken by the magnitude of the environment 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 assortment strategy.

Section 6.3. To isolate detection performance, we exclude regret due to estimating post-change preferences. Instead of empirical averages, we compute regret using the difference in *expected* revenue under the oracle and the implemented policy. For false alarms, when the change is detected prematurely, we compute the one-period regret as $r(S^*(\nu^1), \nu^1) - r(S^*(\nu^\tau), \nu^1)$. Note that the scenario in which a false alarm leads the retailer to learn the pre-change (rather than post-change) preferences is not captured in our current regret calculation. However, we do address this possibility in both our theoretical results and the numerical examples presented throughout the paper.

All policies (based on a change detection approach) introduced in this paper partition the horizon into sub-segments dedicated to either exploration or exploitation. The optimal size of these segments depends on the similarity between the pre- and post-change preferences. Loosely speaking, smaller changes require longer segments of exploration 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 exploration of order at least $\mathcal{O}(10^6)$.

For the active-monitoring-then-learn policy (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 exploitation and $D_2 \log T$ for exploration, with $D_1 = 100$ and $D_2 = 5{,}000$. We construct the set of test assortments using the partitioning approach from Section 5.3.3. The same principle applies for the passive-monitoring-then-learn policy (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 6, we adopt the same sub-segment sizes as for the unknown-preferences setting. Moreover, we use Algorithm 5 with initial assortment $S^*(\nu^1)$ and z_{REV} to obtain the test assortment used as an input of Algorithm 4.

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.

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

Electronic Companion

(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' preference $F^{(\mathbb{N})} \in \mathcal{F}$, 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.

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-and-learn policy, 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 changes in customers' preferences which belong to $\mathcal{F}$. 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:

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

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:

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

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 $\{N\}$:

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

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 $\mathcal{K}(F_a, F_b\,;\, S) = 0$. Indeed, conditional on $S \neq \{N\}$, the distributions F_a and F_b coincide. However, conditional on the assortment $S = \{N\}$, we have that $\mathcal{K}(F_a, F_b\,;\, S) \neq 0$ 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:

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

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 customers' preferences $F, G \in \mathcal{M}^{'}$ 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:

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

where similar observation remains valid whenever the preferences F are replaced by G.

Then, 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 F and G. Formally:

$$\begin{aligned}
\mathcal{K}\big(\mathbb{P}_F^{\pi,\mathcal{F}_j}, \mathbb{P}_G^{\pi,\mathcal{F}_j}\big) &:= \mathbb{E}_{\mathbb{P}_F^{\pi,\mathcal{F}_j}}\big[\log\big(\frac{\mathbb{P}_F^{\pi,\mathcal{F}_j}\big[Z\big]}{\mathbb{P}_G^{\pi,\mathcal{F}_j}\big[Z\big]}\big)\big] \\
&= \mathbb{E}_{\mathbb{P}_F^{\pi,\mathcal{F}_j}}\big[\log\big(\frac{\prod_{t\in\mathcal{F}_j} F^t(Z^t \mid \psi_t)}{\prod_{t\in\mathcal{F}_j} G^t(Z^t \mid \psi_t)}\big)\big] \\
&= \mathbb{E}_{\mathbb{P}_F^{\pi,\mathcal{F}_j}}\big[\sum_{t\in\mathcal{F}_j} \log\big(\frac{F^t(Z^t \mid \psi_t)}{G^t(Z^t \mid \psi_t)}\big)\big] \\
&= \sum_{t\in\mathcal{F}_j} \mathbb{E}_{F^t}\big[\log\big(\frac{F^t(Z^t \mid \psi_t)}{G^t(Z^t \mid \psi_t)}\big)\big] \\
&\overset{(a)}{=} \sum_{t\in\mathcal{F}_j} \mathbb{E}_{F^t}\big[\log\big(\frac{F^t(Z^t \mid \{N\})}{G^t(Z^t \mid \{N\})}\big)\mathbf{1}\big(\psi_t = \{N\}\big)\big]
\end{aligned}$$

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

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

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:

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

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:

$$\begin{aligned}
J^*(F^{(\mathbb{N})},T)-J^{\pi}(F^{(\mathbb{N})},T)\geq&\sum_{j\in[\tilde{T}-1]}\big[\frac{1}{2}\big(\Delta r(\{1\},F_a)-\sum_{t\in\mathcal{F}_j}r(\psi_t,F_a)\big)\\
&\qquad+\frac{1}{2}\big(\Delta r(\{N\},F_b)-\sum_{t\in\mathcal{F}_j}r(\psi_t,F_b)\big)\big]\\
\overset{(a)}{\geq}&\sum_{j\in[\tilde{T}-1]}\frac{\zeta}{2}\big(\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_a}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t\neq\{1\}\big)\big]+\mathbb{E}_{\mathbb{P}^{\pi,\mathcal{F}_j}_{\mathbf{F}_b}}\big[\sum_{t\in\mathcal{F}_j}\mathbf{1}\big(\psi_t\neq\{N\}\big)\big]\big)\\
\geq&\frac{1}{16}\exp\left(-1\right)\zeta(\tilde{T}-1)\Delta\geq\frac{\sqrt{2}-1}{(16)^2\sqrt{2}}\exp\left(-1\right)T^{\frac{3}{4}}M_T^{\frac{1}{4}},
\end{aligned}$$

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-and-learn policy, 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:

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

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:

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

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 $\sum\limits_{j=1}^{\tilde{T}-1} M_j \leq \mathcal{M}(\mathcal{F},T)$.

Let $t_j \in \text{argmin}\{p_0(S^*(F^t), F^t) : t \in \mathcal{F}_j\}$. 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:

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

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 provide proofs for both the lower bound on achievable performance and an upper bound on the regret achieved by Algorithm 1 in settings in which the change is passively undetectable and the retailer has no information about it (except that the change is abrupt). Next, in Section E.C.2.2, 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 magnitude. In Section E.C.2.3, we focus on cases where the change is detectable using information 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.

E.C.2.1 Proofs for Section 5.2

For $T \geq 2$ and $\Delta \in [T]$, we define $\tilde{T}-1 := \lceil T/\Delta \rceil - 1$ as the number of sub-segments $\mathcal{F}_1, \ldots, \mathcal{F}_{\tilde{T}-1}$, 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 $F^{(\mathbb{N})} \equiv \big(F^t \,:\, t \in \mathbb{N}\big)$ 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 $\tau = \ell_j$.

The lower bound that any admissible policy must incur, as stated in Proposition 3, is closely related to the result in Proposition 7, in which the post-change preferences are assumed to be known by the retailer. Specifically, the arguments that we use in the proofs of Lemma 4 and Lemma 5 can be adapted to establish Lemma 2 and Lemma 3. Therefore, to maintain brevity, the proofs of Lemma 2 and 3 bellow are omitted. Finally, we conclude the section with the proof of Proposition 2, which essentially provides an upper bound on the regret achieved by the restart-and-learn policy.

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

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

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

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

Lemma 3. *Assume that there exist an admissible policy $\pi \in \mathcal{P}$ and a constant $\beta > 0$ that satisfy:*

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

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

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

Both lemmas essentially establish a lower bound on the regret that any admissible policy must achieve by classifying policies into two categories: those that explore sufficiently within each possible customer sub-segment and those that do not. Accordingly, we obtain the following proposition.

Proof of Proposition 1. For $T \geq 2$, we define $\Delta = \lceil T^{1/2} \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1/2} \rceil$. Consequently, we have $\Delta \geq T^{1/2} \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1/2}$ and $\tilde{T} - 1 \geq T^{1/2} \cdot \mathcal{M}(\mathcal{F}_U, T)^{1/2} - 1$. Next, we fix $\beta = 1$.

For any policy, we apply either Lemma 2 or Lemma 3, depending on whether the policy explores sufficiently or not. Consequently, the following lower bound for the regret holds:

$$\begin{aligned} \sup_{F^{(\mathbb{N})} \in \mathcal{F}_U} \left\{ J^*(F^{(\mathbb{N})}, T) - J^{\pi}(F^{(\mathbb{N})}, T) \right\} &\geq \min\left\{ C_1\Delta, C_2(\tilde{T} - 1) \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1} \right\} \\ &\geq C(\gamma)\big(T^{1/2} \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1/2} - 1\big), \end{aligned}$$

where $C_1 \equiv C_1(\gamma, 1)$ and $C_2 \equiv C_2(\gamma, 1)$ are the constants obtained from Lemma 2 and Lemma 3, respectively. Moreover, $C(\gamma) = \min\left\{ C_1(\gamma, 1), C_2(\gamma, 1) \right\}$ and the inequality holds for all $T \geq 2$. ■

Next, we derive an upper bound on the regret achieved by the assortment strategy from Algorithm 1. Formally, we provide the proof of Proposition 2.

Proof of Proposition 2. Let F^1 be such that $F^{(\mathbb{N})} \in \mathcal{F}_U$ so that preferences are indistinguishable from $S^*(F^1)$. Loosely speaking, since the environment is aware that the restart-and-learn policy, described in Algorithm 1, always explores assortments, it should never change the customers' preferences. In that case, the policy would repeatedly explore, driving in turn the regret upwards. We formalize this intuition afterwards.

Observe that the minimax regret that is achieved by any policy in the stationary setting does not depend on F^1 or F^{τ} (the pre- and post-change preferences). Let $\mathcal{A} \in \mathcal{P}$ be a subroutine that is

used to learn the customers' preferences in the stationary regime. For $T \geq 2$, we define the segment size used in Algorithm 1 as $\Delta = \lceil T \cdot \mathcal{M}(\mathcal{F}_U, T) \cdot M^{-1} \rceil \leq T$. Accordingly, we have that:

$$\tilde{T} := \lceil T/\Delta \rceil \leq T/\Delta + 1 \leq 2M \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1}.$$

Let ψ_t be assortment decision from policy $\mathcal{A}$ at time t. Then, we derive an upper bound for the difference between the expected revenue obtained by the oracle and that achieved by our policy:

$$\begin{aligned} J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T) &= \sum_{t=1}^{T} \big(r(S^*(F^t), F^t) - r(\psi_t, F^t)\big) \\ &= \sum_{t=1}^{T} \big(r(S^*(F^1), F^1) - r(\psi_t, F^1)\big)\mathbf{1}\big(t < \tau\big) \\ &\quad + \sum_{t=1}^{T} \big(r(S^*(F^\tau), F^\tau) - r(\psi_t, F^\tau)\big)\mathbf{1}\big(t \geq \tau\big) \\ &\leq 2(\tilde{T} - 1) \cdot \mathcal{R}^{\mathcal{A}}\left(\mathcal{F}_b(F^1), \Delta\right). \end{aligned}$$

Accordingly, we obtain the following upper bound on the regret:

$$\sup_{F^{(\mathbb{N})} \in \mathcal{F}_U(F^1)} \left\{J^*(F^{(\mathbb{N})}, T) - J^\pi(F^{(\mathbb{N})}, T)\right\} \leq 4M \cdot \mathcal{M}(\mathcal{F}_U, T)^{-1} \cdot \mathcal{R}^{\mathcal{A}}\left(\mathcal{F}_b(F^1), \Delta\right).$$

which, in turn, concludes the proof. ■

E.C.2.2 Proofs for Section 5.3

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 3. We then derive an upper bound on the regret achieved by the assortment strategy described in Algorithm 2, as formalized in Proposition 4.

Proof of Proposition 3. The proof closely follows the argument the proof of Proposition 7, 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 4. 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 $\pi \equiv \pi(\kappa, F^1, \mathcal{E}, \mathcal{A})$ defined by 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:

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

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 corre-

sponds 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 $\mathcal{A}$ in the stationary setting. Formally, the following inequality holds:

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

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:

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

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:

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

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:

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

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

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 $\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\tau - \hat{k})^+\big]$:

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\tau - \hat{k})^+\big] &\leq |\mathcal{E}|\Delta_e \frac{j^*(j^*-1)}{2} \max\{q_{f,j} \,:\, j \in [\tilde{T}-1]\} \\ &\overset{(a)}{\leq} \frac{T^2}{2|\mathcal{E}|\Delta_e} K(4\kappa^{-2}(\log T)+1)T^{-2} = K\Big(\frac{1}{2|\mathcal{E}|} + \frac{\kappa^2}{8|\mathcal{E}|\log 2}\Big),\end{aligned}$$

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 the following, 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:

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k} - \tau)^+\big] = \sum_{u=0}^{T-\tau+1} \mathbb{P}_\tau^\pi\big[\hat{k} \geq \tau + u\big] &\leq \sum_{j=j^*}^{\tilde{T}-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \mathbb{P}_\tau^\pi\big[\bigcap_{m=j^*}^{j} \big\{\hat{k} \neq \ell_m + 1\big\}\big] \\ &\overset{(a)}{\leq} \sum_{j=j^*+2}^{\tilde{T}-1} \sum_{\ell=\ell_j}^{\ell_{j+1}-1} \mathbb{P}_\tau^\pi\big[\bigcap_{m=j^*+2}^{j} \big\{\hat{k} \neq \ell_m + 1\big\}\big] + 2|\mathcal{E}|\Delta_e \\ &\overset{(b)}{=} |\mathcal{E}|\Delta_e\big(2 + \sum_{j=j^*+2}^{\tilde{T}-1} (q_{d,j^*+2})^{j-j^*-1}\big) \\ &= |\mathcal{E}|\Delta_e\big(2 + q_{d,j^*+2}\frac{1-(q_{d,j^*+2})^{\tilde{T}-j^*}}{1-q_{d,j^*+2}}\big) \\ &\leq \frac{2|\mathcal{E}|\Delta_e}{1-q_{d,j^*+2}} = \frac{8|\mathcal{E}|\log(T)/\kappa^2}{1-q_{d,j^*+2}},\end{aligned}$$

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 assortment $S \in \mathcal{E}$:

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

where $Z^t(S)$ represent 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:

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

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:

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

Step 6. Least but not last, we derive an upper bound for the term $\sum_{S\in\mathcal{E}}\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[\sum_{t=1}^{\hat{k}}\mathbf{1}\big(\psi_t = S\big)\big]$. 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 $T \geq 2$:

$$\sum_{S\in\mathcal{E}} \mathbb{E}_{\mathbb{P}_\tau^\pi}\big[\sum_{t=1}^{\hat{k}} \mathbf{1}\big(\psi_t = S\big)\big] \leq \sum_{S\in\mathcal{E}} (\tilde{T}-1)\Delta_e \leq \sum_{S\in\mathcal{E}} \frac{T}{\Delta_o + |\mathcal{E}|\Delta_e}\Delta_e$$
$$= |\mathcal{E}| \frac{4T(\log T)/(\kappa^2)}{\sqrt{T}/\kappa^2 + 4\,|\mathcal{E}|\,(\log T)/(\kappa^2)} \leq 4|\mathcal{E}|\sqrt{T}\log T.$$

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.2.3 Proofs for Section 5.4

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 5, and derive an upper bound on the regret achieved by Algorithm 3 as outlined in proof of Proposition 6.

Proof of Proposition 5. The proof closely follows the one for Proposition 9 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 9 can be replaced by ε (see the proof of Proposition 9 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 6. 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 $j \in \{0, \dots, \tilde{T} - 1\}$, 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 learning algorithm (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 learning algorithm. 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:

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

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 $q_{f,j} := \mathbb{P}\big[\hat{\Lambda}_{\ell_j} = 1 \mid H_{0,j}\big]$. 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:

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

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:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^\pi_\tau}\big[(\tau - \hat{k})^+\big] &\leq \Delta \frac{j^*(j^* - 1)}{2} q_{f,j} \\
&\leq K\Delta \frac{\tilde{T}(\tilde{T} - 1)}{2} \exp(-\frac{1}{2}\Delta\varepsilon^2) \\
&\leq \Delta^{-1} T^2 K(\Delta + 1) \exp(-\frac{1}{2}\Delta\varepsilon^2) = K(1 + \Delta^{-1}) \leq K(1 + \frac{\varepsilon^2}{4\log(2)}).
\end{aligned}$$

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

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

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^*:

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

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

Hence, the following inequalities hold:

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k}-\tau)^+\big] &\leq 2\Delta\big(1-K(\Delta+1)\exp(-2\Delta(\|p(S^*(F^1),F^1)-p(S^*(F^1),F^\tau)\|_\infty-\frac{\varepsilon}{2})^2)\big)^{-1}\\ &\overset{(a)}{\leq} 2\Delta\big(1-K(\Delta+1)\exp(-2\Delta(\sqrt{\varepsilon/2}-\varepsilon/2)^2)\big)^{-1}\overset{(b)}{\leq} 6\Delta,\end{aligned}$$

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

This section contains the proofs for the theoretical results established in Appendix A, which addresses the case in which the post-change preferences are known by the retailer. Our analysis proceeds in three stages. First, 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 minimum 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.1

This section establishes proofs for Proposition 7 and Proposition 8, about achievable performances and regret of Algorithm 4. Then, we provide the proof of Lemma 1. 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 4 analyzes policies that do not sufficiently explore alternative assortments within at least one sub-segment. Then, Lemma 5 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 4. *Assume that there exist an admissible policy $\pi \in \mathcal{P}$ and a constant $\beta > 0$ that satisfy:*

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

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:

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

Next, let $i_0 \in [\tilde{T}-1]$ be such that $\mathcal{K}(\mathbb{P}^\pi_{\ell_{i_0+1}}, \mathbb{P}^\pi_{\ell_{i_0}}) \le \beta$. 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:

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

In other words, the following inequality holds:

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

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:

$$\sup_{k \in \{i_0, i_0+1\}} \mathbb{E}_{\mathbb{P}^\pi_{\ell_k}}\big[J^*(F^{(\mathbb{N})}, T) - \mathcal{J}^\pi(F^{(\mathbb{N})}, T)\big] \le \tilde{C}\Delta. \tag{E.C.33}$$

Next, consider the following decision rule:

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

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:

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

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 8, and (c) uses the bound from equation (E.C.33).

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:

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

$$= \gamma\big(\Delta - \sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1} \mathbf{1}\left[\|S^*(F^1) - \psi_t(\mathcal{H}_{t-1})\|_1 \geq 1\right]\big)$$

$$= \gamma\big(\Delta - \Delta/2\big) = \gamma\Delta/2.$$

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

$$\begin{aligned}
\mathbb{P}^{\pi}_{\ell_{i_0}}\big[\phi = 0\big] &\leq \mathbb{P}^{\pi}_{\ell_{i_0}}\Big[\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\big(r\left(S^*(F^\tau), F^\tau\right) - r\left(\psi_t(\mathcal{H}_{t-1}), F^\tau\right)\big) \geq \frac{\gamma\Delta}{2}\Big] \\
&\overset{(a)}{\leq} \frac{2}{\gamma\Delta}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{i_0}}}\Big[\sum_{t=\ell_{i_0}}^{\ell_{i_0+1}-1}\big(r\left(S^*(F^\tau), F^\tau\right) - r\left(\psi_t(\mathcal{H}_{t-1}), F^\tau\right)\big)\Big] \\
&\overset{(b)}{\leq} \frac{2}{\gamma\Delta}\mathbb{E}_{\mathbb{P}^{\pi}_{\ell_{i_0}}}\big[J^*(F^{(\mathbb{N})}, T) - \mathcal{J}^{\pi}(F^{(\mathbb{N})}, T)\big] \overset{(c)}{\leq} \frac{2}{\gamma\Delta}\tilde{C}\Delta = \frac{1}{2}\max\big\{\frac{1}{4}\exp(-\beta), \frac{1-\sqrt{\beta/2}}{2}\big\},
\end{aligned}$$

where (a) follows from Markov's inequality (Jacod and Protter 2012), while (b) follows from the optimality of $S^*(F^\tau)$ and Lemma 8. Finally the last inequality (c) follows from (E.C.33).

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:

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

which is a contradiction with (E.C.33). 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 5. *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:

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

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:

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

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 12; see (E.C.4).

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

$$\sup\big\{\big|\log p_i(S, F^1) - \log p_i(S, F^\tau)\big| \,:\, \forall\, i \in S \cup \{0\},\, \forall\, S \in \mathcal{S}\big\} \le \vartheta,$$

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

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

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

We now present the proof of Proposition 7. The proof follows from the observation that for any given $\beta > 0$, the conditions of either Lemma 4 or Lemma 5 must be satisfied. The proposition's conclusion therefore follows directly from the application of the relevant lemma.

Proof of Proposition 7. 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 4 or Lemma 5, depending on whether the policy explores sufficiently or not.

As a result, the following inequality holds:

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

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

We now present the proof for the upper bound on the regret achieved by the active-monitoring-then-optimize policy, as described in Algorithm 4. Formally, we provide the proof of Proposition 8.

Proof of Proposition 8. 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-then-optimize policy, 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:

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

Step 2. We begin by providing an upper bound to $\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\tau - \hat{k})^+\big]$. 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:

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

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. 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 customer $\ell_{j+1} - \Delta_e - 1$ to $\ell_{j+1} - 1$ are available.

Given these purchase decisions, we consider the following two statistical hypothesis:

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

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:

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

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 $\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[(\hat{k} - \tau)^+\big]$. 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:

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

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:

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

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:

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

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 $C := \delta(1 + (4D(\alpha))^{-1} + 6D(\alpha))$, then we conclude the proof. ■

Finally, we prove Lemma 1, which establishes the theoretical guarantees for our proposed procedure to identify test assortments capable of distinguishing between pre- and post-change preferences.

Proof of Lemma 1. Let $\mathcal{T}$ denote Algorithm 5 as in the lemma for given parameters $S^0 \in \mathcal{S}$, $K > 0$, and $z \in \{z_{\text{SEP}}, z_{\text{REV}}\}$. In the worst case, the procedure must examine all k-flip neighborhood of S^0, for $k \in [K]$. Consequently, the algorithm terminates after enumerating at most

$$\sum_{k=1}^{K} \binom{K}{k} = \mathcal{O}(N^K)$$

possible assortments, establishing its worst-case running-time.

Next, let S^* be the assortment returned by the procedure at iteration $k \in [K]$. By construction, we have $z(N_k(S^0)) > 0$. Thus, we obtain $z(\{S^*\}) > 0$, which concludes the proof. ■

E.C.3.2 Proofs for Appendix A.2

Next, we establish a lower bound on achievable performance and an upper bound on the regret attained by Algorithm 6. Specifically, we present proofs for Proposition 9 and 10, which rely on two technical results, namely, Lemmas 6 and 7 that are proved in (E.C.3.3).

Proof of Proposition 9. 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 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 8). For an arbitrary $\eta > 0$, we define:

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

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

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

and, for $i \ge 1$:

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

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

Lemma 6. *For any $\omega \in B_\eta$, we have: $0 \leq \hat{k}^* - \tau \leq 2j_0$.*

By leveraging Lemma 6, we establish the following chain of set inclusions:

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

Consequently, the following inequality is valid:

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

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

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

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

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

Lemma 7. *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:*

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

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 7. Then, we derive the following sequence of inequalities:

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

We can now establish a sequence of inequalities that lead to our main result. First, applying our previous findings and Lemma 7, we obtain:

$$\sup_{1\leq\tau\leq T} \mathbb{P}_\tau^\pi\big[B_\eta^c\big] \geq \sup_{1\leq\tau\leq T} \mathbb{P}_\tau^\pi\big[\big|\hat{k} - \tau\big| > 2j_0\big] \geq \sup_{1\leq\tau\leq T} \mathbb{P}_\tau^\pi\big[\big|\hat{k} - \tau\big| > \tilde{C}\log T\big] \geq \alpha.$$

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:

$$\sup_{F^{(\mathbb{N})}\in\mathcal{F}(F^1,F^\tau)} \big\{J^*(F^{(\mathbb{N})},T) - J^\pi(F^{(\mathbb{N})},T)\big\} \geq \sup_{1\leq\tau\leq T} \eta\mathbb{P}_\tau^\pi\big[B_\eta^c\big] \geq \sup_{1\leq\tau\leq T} \alpha\big(C_1\log T - \gamma\big)^+ \overset{(a)}{=} C\log T,$$

where equality (a) holds for all $T \geq \exp\big(\frac{4}{C}\big)$ 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-then-optimize policy. Recall that this policy is formally defined in Algorithm 6.

Proof of Proposition 10. Assume that $T \geq 2$, F^1 and F^τ are such that the preferences $F^{(\mathbb{N})}$ they induce belong to $\mathcal{F}_D$. Define $\alpha \equiv (\alpha_{\mathrm{I}}, \alpha_{\mathrm{II}})$ 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:

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

We define the customer batch size, denoted by $\Delta := D\log T$, as specified in Algorithm 6. 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 $\ell_{j^*} < \tau \leq \ell_{j^*+1}$, 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 6. 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:

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

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:

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

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

Then, we derive the appropriate upper bound for the regret of policy π.

Step 2. We proceed to analyze the expected detection advance $\mathbb{E}_{\mathbb{P}_\tau^\pi}\big[\big(\tau-\hat{k}\big)^+\big]$ 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:

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

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 the fact that $\ell_{j^*}<\tau\leq\ell_{j^*+1}$.

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:

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

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, we have that:

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

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. Therefore, we conclude the proof. ■

E.C.3.3 Proofs of Lemmas 6 and 7

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 6 and 7.

Proof of Lemma 6. 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:

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

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:

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

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:

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

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 7. 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:

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

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

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

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 $Z := \left(Z^1, \ldots, Z^T\right)$ 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:

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

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:

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

$$= \int_\Omega \prod_{t=\ell_j}^{\ell_{j+1}-1} \frac{F^1\big(Z^t \mid \psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)}{F^\tau\big(Z^t \mid \psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)} V(\omega)$$

$$\cdot \prod_{t=1}^{\ell_j-1} F^1\big(Z^t \mid \psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big) \prod_{t=\ell_j}^{T} F^\tau\big(Z^t \mid \psi_t\big(\mathcal{H}_{t-1}(\omega)\big)\big)\, d\omega$$

$$= \mathbb{E}_{\mathbb{P}^\pi_{\ell_{j+1}}}\big[V\big],$$

where we use a change of measure type of argument, together with Lemma 9.

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

$$\mathbb{E}_{\mathbb{P}^\pi_{\ell_j}}\big[e^{\tilde{Z}_j}\mathbf{1}\big(|\hat{k}-\ell_j| \leq \Delta/3\big)\big] \geq \mathbb{E}_{\mathbb{P}^\pi_{\ell_j}}\big[e^{-\Delta\vartheta}\mathbf{1}\big(|\hat{k}-\ell_j| \leq \Delta/3\big)\big]$$

$$\geq \mathbb{E}_{\mathbb{P}^\pi_{\ell_j}}\big[e^{-\left(1+C\log T\right)\vartheta}\mathbf{1}\big(|\hat{k}-\ell_j| \leq \Delta/3\big)\big] = \frac{e^{-\vartheta}}{T^{C\vartheta}}\mathbb{E}_{\mathbb{P}^\pi_{\ell_j}}\big[\mathbf{1}\big(|\hat{k}-\ell_j| \leq \Delta/3\big)\big].$$

Consequently, we derive the following sequence of inequalities:

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

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

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

$$e^{-\vartheta}T^{-C\vartheta}(\tilde{T}-1) \geq e^{-\vartheta}T^{-C\vartheta}\frac{\tilde{T}}{2}$$

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

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

Consequently, we obtain the following inequalities:

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

which is clearly a contradiction. Therefore, we conclude:

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

That is, for all $T \geq n_1$, we have:

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

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 $\mathbb{P}_{\tau_1}\big[|\hat{k}-\tau_1| = 0\big] \geq \frac{1}{2}$. Next, observe that:

$$\mathbb{P}^\pi_{\tau_2}\big[|\hat{k}-\tau_1| = 0\big] = \mathbb{E}^\pi_{\tau_2}\big[\mathbf{1}\big(|\hat{k}-\tau_1| = 0\big)\big]$$

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

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

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

Therefore, we obtain the following inequalities:

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

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

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

Therefore, we obtain the following inequality:

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

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

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

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

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

where $\tilde{C} := \min\big\{C/3, 1/n_1\big\}$ and $\alpha := \min\big\{1-\beta, e^{-\vartheta}/2\big\}$. 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. We provide formal statements and proofs of the key lemmas and corollary highlighted in Table E.C.1.

Lemma	Corollary
8, 9, 10, 11, 12	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*:

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

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 8. *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}_{\tau}^{\pi}}\big[\sum_{t=1}^{T}\sum_{i\in\psi_t} w_i \mathbf{1}(i_t = i)\big] = \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\sum_{t=1}^{\tau-1}\sum_{i\in\psi_t} w_i \mathbf{1}(i_t = i)\big] + \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i \mathbf{1}(i_t = i)\big], \\
&= \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\sum_{t=1}^{\tau-1}\sum_{i\in\psi_t} w_i \mathbf{1}\big(i_t = i \in \psi_t, U_{i_t}^{a} > U_{j}^{a},\ \forall\, j \in \psi_t \cup \{0\} \setminus \{i_t\}\big)\big] \\
&\quad + \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i \mathbf{1}\big(i_t = i \in \psi_t, U_{i_t}^{b} > U_{j}^{b},\ \forall\, j \in \psi_t \cup \{0\} \setminus \{i_t\}\big),\big] \\
&\overset{(a)}{=} \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\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_{i_t}^{a} > U_{j}^{a},\ \forall\, j \in \psi_t \cup \{0\} \setminus \{i_t\}\big) \mid \mathcal{H}_{t-1}\big]\big] \\
&\quad + \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i \mathbf{1}\big(i_t = i \in \psi_t, U_{i_t}^{b} > U_{j}^{b},\ \forall\, j \in \psi_t \cup \{0\} \setminus \{i_t\}\big) \mid \mathcal{H}_{t-1}\big]\big] \\
&\overset{(b)}{=} \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\sum_{t=1}^{\tau-1}\sum_{i\in\psi_t} w_i \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\mathbf{1}\big(i_t = i \in \psi_t, U_{i_t}^{a} > U_{j}^{a},\ \forall\, j \in \psi_t \cup \{0\} \setminus \{i_t\}\big) \mid \mathcal{H}_{t-1}\big]\big] \\
&\quad + \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\sum_{t=\tau}^{T}\sum_{i\in\psi_t} w_i \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\big[\mathbf{1}\big(i_t = i \in \psi_t, U_{i_t}^{b} > U_{j}^{b},\ \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}_{\tau}^{\pi}}\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}_{\tau}^{\pi}}\big[\sum_{i\in\psi_t} w_i p_i\left(\psi_t, F^{\tau}\right)\big] = \mathbb{E}_{\mathbb{P}_{\tau}^{\pi}}\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 8, 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 $T \geq 2$:*

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

Proof. The proof immediately follows from Lemma 8. ■

Next, we formally define and derive a closed-form expression for the probability of purchase over a finite sequence of T customers, denoted by $\mathbb{P}_{\ell}^{\pi}$. 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 9. *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:*

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

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

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

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:

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

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 $\mathbb{P}^{\pi}_{\ell}\big[Z = z\big] = 0$, which concludes the proof. ■

The distribution $\mathbb{P}^{\pi}_{\ell}$ 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 10 provides a closed-form expression for this divergence.

Lemma 10. *For $\ell_j, \ell_{j+1} \in \{1, \ldots, T\}$, let $\mathcal{K}(\mathbb{P}^{\pi}_{\ell_{j+1}}, \mathbb{P}^{\pi}_{\ell_j})$ denote the KL divergence between the two probability measures $\mathbb{P}^{\pi}_{\ell_j}$ and $\mathbb{P}^{\pi}_{\ell_{j+1}}$. Then, we have that:*

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

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 9 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 11. *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 12. *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 $t \in \big[T\big]$. By Lemma 10, 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:

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

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.