Last-iterate Convergence in No-regret Learning: Games with Reference Effects Under Logit Demand*

Mengzi Amy Guo1, Donghao Ying2, Javad Lavaei3, Zuo-Jun Max Shen4
Department of Industrial Engineering and Operations Research, University of California, Berkeley
1 mengzi.guo@berkeley.edu, 2 donghaoy@berkeley.edu, 3 lavaei@berkeley.edu, 4 maxshen@berkeley.edu

This work is dedicated to the algorithm design in a competitive framework, with the primary goal of learning a stable equilibrium. We consider a realistic setting where two firms engage in a multi-period price competition within an opaque marketplace—each firm lacks information about its competitor—under the influence of reference effects. Consumers assess their willingness to pay by comparing the current price against a reference point formed from historical prices, and their demand follows the multinomial logit (MNL) choice model. We use the notion of stationary Nash equilibrium (SNE), defined as the fixed point of the equilibrium pricing policy, to simultaneously capture the long-run market equilibrium and stability. With the loss-neutral reference effect, we propose the online projected gradient ascent (OPGA) algorithm, where the firms adjust prices using the first-order derivatives of their log-revenues obtained from the market feedback mechanism. Despite the absence of typical properties required for the convergence of online games, such as strong monotonicity and variational stability, we demonstrate that under diminishing step-sizes, the price and reference price paths generated by the OPGA attain last-iterate convergence to the unique SNE, and thereby guarantee the no-regret learning and market stability. Moreover, with appropriate step-sizes, we prove that this algorithm exhibits a convergence rate of $O(1/t)$ and achieves a dynamic regret of $O(\sqrt{T})$ over $T$ periods. The inherent asymmetry nature of reference effects motivates explorations beyond the loss-neutral scenario. When loss-averse reference effects are introduced, we propose the conservative-OPGA (C-OPGA) algorithm to adapt to the non-smooth revenue functions and demonstrate that the price and reference price achieve last-iterate asymptotic convergence to the set of SNEs. By contrast, the presence of any gain-seeking product precludes the existence of an SNE, consequently eliminating the possibility of simultaneously attaining market equilibrium and stability.

Key words: last-iterate convergence, price competition, reference effect, multinomial logit model

History:

* We note that the algorithm and its convergence for the case of loss-neutral reference effects (i.e., Theorems 1 and 2) were proposed in our conference proceeding paper (Guo et al. 2023). In this work, however, we extend the analysis to consider the more general setting with potentially asymmetric reference effects. Additionally, we prove new theoretical results regarding the regret bound, which is not included in Guo et al. (2023).
1. Introduction

The memory-based reference effect is a well-established strategic consumer behavior in marketing and economics literature, which refers to the phenomenon that consumers shape their price expectations, known as reference prices, based on their past encounters and then use them to judge the current price (see Mazumdar et al. (2005) for a review). A substantial body of empirical research has shown that the current demand is significantly influenced by the historical prices through reference effects (see, e.g., Krishnamurthi et al. (1992), Hardie et al. (1993), Greenleaf (1995), Briesch et al. (1997), Kalyanaram and Winer (1995)). Driven by the ubiquitous evidence, many studies in the operations research community have investigated effective pricing strategies in the presence of reference effects (see, e.g., Popescu and Wu (2007), Fibich et al. (2003), Chen et al. (2017), Chen and Nasiry (2020), Jiang et al. (2022)). While the aforementioned works all focus on scenarios of a monopolistic seller, little is known about the role of reference effects within a competitive framework, especially considering its practical importance. This problem becomes even more pronounced with the surge of e-commerce, where a plethora of accessible information encourages consumers to form reference prices and draw comparisons between various alternatives and retailers.\(^1\)

Moreover, we observe that the intricacy of pricing in a competitive market is further compounded by the issue of opacity, where firms are cautious about revealing information to their rivals. Although the rise of digital markets greatly accelerates data transmission and promotes information transparency, which is generally considered beneficial because of the improvement in marketplace efficiency (Tapscott and Ticoll 2003), many firms remain hesitant to fully embrace such transparency for fear of losing their informational advantages. This concern is well-founded in the literature: for example, Bloomfield and O’Hara (2000) demonstrate that dealers with lower transparency levels generate higher profits than their more transparent counterparts, and Wilson (2000) highlights that in reverse auctions, transparency typically enables buyers to drive the price down to the firm’s marginal cost. Hence, as recommended by Granados and Gupta (2013), companies should focus on the strategic and selective disclosure of information to enhance their competitive edge, rather than pursuing complete transparency.

Inspired by these real-world practices, in this article, we study the duopoly competition with reference effects in an opaque market, where each firm has access to its own information but does not possess any knowledge about its competitor, including their price, reference price, and demand. The consumer demand follows the multinomial logit (MNL) choice model, which naturally captures the interactions among substitutes in a multi-product setting. Furthermore, given the intertemporal

\(^1\)As e-commerce and online selling continue to surge rapidly, price tracking becomes increasingly convenient, making historical prices more readily available to consumers. An example of this is camelcamelcamel.com, a free price tracking service for Amazon.com, one of the leading online retailers in the US.
characteristic of the memory-based reference effect, we consider the game in a dynamic/multi-period framework, i.e., the firms engage in repeated competitions with consecutive periods linked by reference price updates. In this problem, a natural question to ask is whether the firms can achieve last-iterate convergence to some notion of stable equilibrium in the long run by employing no-regret learning algorithms to sequentially set their prices within this opaque marketplace. We note that an algorithm is classified to be no-regret learning if the regret associated with the sequence of online actions produced by the algorithm, when compared to the best fixed action in hindsight, grows sub-linearly with the total number of episodes. However, it is important to mention that being no-regret does not guarantee the convergence to an equilibrium by any means (Mertikopoulos et al. 2018). In the online learning literature, guaranteeing last-iterate convergence is typically more challenging than merely demonstrating a sublinear regret, as achieving the former inherently proves the latter.

A rich literature on online games with incomplete information targets the problem of finding no-regret algorithms that can direct agents toward a Nash equilibrium, a stable state at which the agents have no incentive to modify their actions (see, e.g., Bravo et al. (2018), Mertikopoulos and Zhou (2019), Lin et al. (2020)). Yet, the Nash equilibrium alone is inadequate in determining the stable state for our problem of interest, given the evolution of reference prices. For instance, even if an equilibrium price is reached in one period, the reference price update makes it highly probable that the firms will deviate from this equilibrium in subsequent periods. As a result, to jointly capture the market equilibrium and stability, we consider the concept of stationary Nash equilibrium (SNE), defined as the fixed point of the equilibrium pricing policy. More precisely, at an SNE, each firm’s Nash equilibrium price with respect to its single-period revenue is equal to the given reference price, which is formally stated in Definition 2. Consequently, the firms have no incentive to deviate once their prices and reference prices converge to the SNE, which ensures that the prices and the corresponding demands remain constant in competition for upcoming periods. Attaining this long-term market stability is particularly appealing to firms, as a stable market creates favorable conditions for implementing efficient planning strategies and facilitating upstream operations in supply chain management (Caves and Porter 1978). In contrast, fluctuating demands necessitate more carefully designed strategies to ensure effective logistics management (Song and Zipkin 1993, Aviv and Federgruen 2001, Trehanne and Sox 2002, Liu et al. 2022).

The concept of SNE has also been investigated by Golrezaei et al. (2020) and Guo and Shen (2023). Particularly, Golrezaei et al. (2020) study the long-run market behavior in a duopoly competition with linear demand and a common reference price for both products. Their results apply exclusively to the symmetric reference effect, i.e., consumers are equally sensitive toward surcharges and discounts, and do not readily generalize to the asymmetric case. Nevertheless,
exploring beyond the symmetric scenario is essential given the intrinsic asymmetry in reference effects, which has its theoretical foundation in prospect theory (Kahneman and Tversky 1979) and is further justified by abundant empirical evidence (see, e.g., Krishnamurthi et al. (1992), Hardie et al. (1993), Kalyanaram and Little (1994), Kalyanaram and Winer (1995), Greenleaf (1995), Bell and Lattin (2000), Slonim and Garbarino (2009)). Alternatively, Guo and Shen (2023) accommodate asymmetric reference effects and adopt the MNL demand with product-specific reference price formulation. Yet, their emphasis diverges significantly from ours: they concentrate on deriving the properties of the equilibrium in a complete information setting, which lacks the learning components that are central to our study. By contrast, this work addresses both the partial information setting and asymmetric reference effects, where we shed light on the market equilibrium and stability through the use of no-regret learning algorithms, i.e., Online Projected Gradient Ascent (OPGA) and Conservative Online Projected Gradient Ascent (C-OPGA). Precisely, both the OPGA and the C-OPGA algorithms are fed with only the firm’s own last-period price and demand as input. That is to say, the implementation of these algorithms does not require competitive intelligence from rival firms, nor does it rely on knowledge of the firm’s own reference price or the reference price update mechanism.

1.1. Contributions
Motivated by real-world retail environments, we introduce a competitive framework with reference effects in an opaque market setup. The intertemporal property of the reference price naturally gives rise to a multi-period competition, where we propose gradient-based learning algorithms that help firms achieve a stable equilibrium over time without requiring any information from their competitors. To the best of our knowledge, this is the first work with a focus on finding a stable equilibrium in such a non-transparent market environment with the accommodation of asymmetric reference effects.

1. Symmetric Reference Effects. In the loss-neutral scenario, we propose the OPGA algorithm, where each firm adjusts its posted price using the first-order derivative of its log-revenue. When the firms execute the OPGA with diminishing step-sizes, we show that their prices and reference prices achieve last-iterate convergence to the unique SNE, thereby assuring no-regret learning and leading to the long-run market equilibrium and stability. Furthermore, when the step-sizes decrease appropriately in the order of $\Theta(1/t)$, we demonstrate that the prices and reference prices converge at a rate of $O(1/t)$ and achieve a dynamic regret of $O(\sqrt{T})$ over $T$ periods.

2. Asymmetric Reference Effects. In the loss-averse scenario, the non-smooth revenue function and the non-convex set of SNEs (where SNE is not unique in the loss-averse scenario) pose significant challenges. To navigate this, we develop the C-OPGA algorithm by introducing a conservative
mechanism in the learning process. Specifically, we propose the concept of virtual gradient to help with the price updates and design a novel metric to evaluate the convergence. Under diminishing step-sizes, we show that the price and reference price paths generated by the C-OPGA achieve last-iterate asymptotic convergence to the set of SNEs. By contrast, an SNE never exists in the presence of any gain-seeking product, thereby excluding the possibility of simultaneously achieving equilibrium and market stability.

3. Convergence Analysis. From the optimization perspective, the analyses of OPGA and C-OPGA are related to the study of online games with gradient feedback and discrete nonlinear systems. However, the incorporation of the MNL model renders our problem lacking favorable properties typically required in the literature, such as strong monotonicity, variational stability, and concavity of the reward function, making the existing theories inapplicable. Additionally, unlike standard online games or nonlinear systems, the dynamic state of reference price and the non-smoothness introduced by the potentially asymmetric reference effects further perplex the convergence analysis. To address these issues in establishing the convergence of the OPGA and the C-OPGA algorithms, we adopt a two-step analysis based on the proximity of the price to the SNE and develop novel techniques that exploit both the global and local properties of the MNL-based objectives. Finally, we show that our algorithms are robust in the sense that the analyses can also be adapted to the inexact gradient oracle, under which the price and reference price would converge to the neighborhood of the SNE.

4. Managerial Insights. The practical importance of this study lies in attaining the long-term market equilibrium and stability for a multi-period game while safeguarding firms’ confidentiality, where the intertemporal aspect of the game is naturally prompted by the key consumer behavior—reference effects. Considering the prevalent concern of firms eager to collaborate yet reluctant to disclose their information to rivals, our proposed the OPGA and the C-OPGA algorithms offer an excellent solution by guiding firms toward the SNE as if they possess complete information.

1.2. Organization

The rest of the paper is structured as follows. In Section 2, we conduct a literature review on topics pertinent to our study. Section 3 unfolds the modeling framework in an opaque market and introduces the notion of SNE for equilibrium and stability characterization. Section 4 is dedicated to the loss-neutral scenario, where we propose the online algorithm OPGA, establish its global convergence to SNE, and derive the dynamic regret bound. The algorithm is then adapted for the loss-averse scenario in Section 5, and its convergence is investigated. In Section 6, we extend the convergence results to the inexact gradient case. Finally, we conclude this work with discussions in Section 7. All formal proofs are documented in the supplemental materials.
2. Related Literature

Our work on the price competition with reference effects in an opaque market is related to several streams of literature: modeling of reference effect, pricing with reference effects in monopolist and competitive markets, and general convergence results for online games.

2.1. Modeling of Reference Effect

The concept of reference effects can be traced back to the adaptation-level theory proposed by Helson (1964), which states that consumers evaluate prices against the level they have adapted to. Meanwhile, the asymmetry nature of reference effects is theoretically grounded in the prospect theory (Kahneman and Tversky 1979), which posits that individuals exhibit the loss-averse behavior, meaning that they weigh losses more heavily than equivalent gains. In addition to the theoretical foundation, the significance of asymmetric reference effects has been substantiated by extensive empirical evidence. Krishnamurthi et al. (1992), Hardie et al. (1993), Kalyanaram and Little (1994), and Kalyanaram and Winer (1995) have calibrated a variety of reference effects models on panel data, where their findings corroborate the presence of loss-averse reference effects. Alongside the loss-aversion, empirical studies (see, e.g., Greenleaf (1995), Mazumdar and Papatla (1995), Briesch et al. (1997), Bell and Lattin (2000)) also report instances of asymmetry in the reverse direction, known as the gain-seeking behavior, where consumers display greater sensitivities to gains than losses. This trend is notably observable in markets dominated by promotion-driven consumers (Kallio et al. 2009) or where products exhibit significant stackability (Slonim and Garbarino 2009). Recognizing the indispensability of this asymmetric attribute, we explore both loss-averse and gain-seeking scenarios in this work.

Extensive research has been dedicated to the formulation of reference effects in the marketing literature, where two mainstream models emerge: memory-based and stimulus-based reference prices (Briesch et al. 1997). The memory-based reference model, also known as the internal reference price, leverages historical prices to form the benchmark (see, e.g., Lattin and Bucklin (1989), Krishnamurthi et al. (1992), Mayhew and Winer (1992), Kalyanaram and Little (1994), Kalyanaram and Winer (1995)). On the contrary, the stimulus-based reference model, or the external reference price, asserts that the price judgment is established at the moment of purchase utilizing current external information such as the prices of substitutable products, rather than drawing on past memories (Lynch Jr and Srull 1982, Biehal and Chakravarti 1983, Hardie et al. 1993). According to the comparative analysis by Briesch et al. (1997), among different reference models, the memory-based formulation that relies on a product’s own historical prices (referred to as PASTBRSP in Briesch et al. (1997)) offers the best fit and strongest predictive power in a multi-product setting. Consequently, our paper adopts this type of reference price model.
2.2. Dynamic Pricing in Monopolist Market with Reference Effect

Recent quantitative analyses on pricing strategies with reference effects predominantly target monopolistic settings. As memory-based reference effect models give rise to the intertemporal nature, the price optimization problem is usually formulated as a dynamic program. Similar to our work, their objectives typically involve determining the long-run market stability under the optimal or heuristic pricing strategies. The classic studies by Fibich et al. (2003) and Popescu and Wu (2007) show that in either a discrete-time or continuous-time framework, the optimal pricing policy converges and leads to market stabilization under both loss-neutral and loss-averse reference effects. More recent research on reference effects primarily concentrates on the piecewise linear demand in single-product contexts and delves into more comprehensive characterizations of myopic and optimal pricing policies (Hu et al. 2016, Chen et al. 2017, Chen and Nasiry 2020). Deviating from the linear demand, Jiang et al. (2022) and Guo et al. (2022) employ the logit demand and analyze the long-term market behaviors under the optimal pricing policy, where the former emphasizes on consumer heterogeneity and the latter innovates in the multi-product setting.

While the aforementioned studies commonly assume that the firm possesses full knowledge of its demand function, another line of research tackles the problem in the context of uncertain demand, where they couple monopolistic price optimization with reference effects and online demand learning (den Boer and Keskin 2022, Ji et al. 2023). Though these works incorporate learning components, our paper differentiates itself in two critical aspects. Firstly, while their objective centers on designing algorithms to boost total revenues from a monopolist perspective, our algorithm targets the simultaneous realization of Nash equilibrium and market stability under a competitive framework. Secondly, the uncertainties to be learned reside in distinct areas. den Boer and Keskin (2022) and Ji et al. (2023) assume that a monopolist seller recognizes the demand structure but needs to estimate sensitivity parameters. By contrast, we suppose that in a competitive environment, a firm knows its own demand but lacks insights into its rivals’ pricing and demand functions.

2.3. Price Competition with Reference Effects

A stream of literature addresses equilibrium in price competition with reference effects under the perfect information setting (Coulter and Krishnamoorthy 2014, Federgruen and Lu 2016, Guo and Shen 2023, Colombo and Labrecciosa 2021). Specifically, the article by Federgruen and Lu (2016) examines single-period price competitions under diverse reference price formulations. Coulter and Krishnamoorthy (2014) expand the scope to a two-period dynamic competition, focusing exclusively on symmetric reference effects. Colombo and Labrecciosa (2021) further extend the game to the multi-stage setting in the continuous-time framework. In a recent study, Guo and Shen (2023) explore the long-run market behavior under the logit demand, assuming a transparent marketplace.
More closely pertinent to our work, Golrezaei et al. (2020) also look into the long-run market stability of the duopoly price competition in an opaque market environment. However, four notable distinctions set our study apart from theirs. The most salient difference lies in our accommodation for asymmetric reference effects. Unlike Golrezaei et al. (2020), which does not account for this asymmetry aspect, we propose a revised version of our original algorithm and validate its convergence to SNE(s) in the loss-averse scenario. Secondly, our preference for the logit demand, as opposed to their linear demand, stems from its capability to accurately represent discrete choices among multiple products and its enhanced performance in the presence of reference effects (Wang 2018, Jiang et al. 2022). Yet, this preference imposes challenges for convergence since a crucial part of the analysis in Golrezaei et al. (2020) hinges on demand linearity (see Golrezaei et al. (2020, Lemma 9.1)), a condition not met by the MNL model. Thirdly, we diverge from Golrezaei et al. (2020) in terms of the reference price formulation. While they presume a uniform reference price for all products, our model incorporates a product-specific reference price, empirically validated for superior performance in Briesch et al. (1997). Lastly, the proof provided by Golrezaei et al. (2020) is contingent upon specific assumptions regarding sensitivity parameters. Our study, however, establishes convergence in a broader sense, without adhering to such stipulations. These adaptations, though enriching the model’s expressiveness and flexibility, render the analysis in Golrezaei et al. (2020) not generalizable to our setting.

2.4. General Convergence Results for Games

Our paper is closely related to the study of general-sum online games with access to the first-order oracle. In this line of research, a typical question is whether learning algorithms can achieve some equilibrium for multiple agents who aim to minimize (resp. maximize) their individual loss (resp. reward) functions. For games with continuous actions, Bravo et al. (2018) and Lin et al. (2021) show the convergence of online mirror descent to the Nash equilibrium in strongly monotone games. Lin et al. (2020) then relax the strong monotonicity assumption and examine the last-iterate convergence for games with unconstrained action sets satisfying the so-called “cocoercive” condition. Mertikopoulos and Staudigl (2017) and Mertikopoulos and Zhou (2019) establish the convergence of the dual averaging method under a more general condition called global variational stability, which encompasses the cocoercive condition as a sub-case. More recently, Golowich et al. (2020), Hsieh et al. (2021) and Hsieh et al. (2022) demonstrate that extra-gradient approaches, like the optimistic gradient method, can achieve faster convergence to equilibrium in monotone and variationally stable games. We point out that in the works cited above, the concavity and smoothness of the individual reward function are invariably required for the convergence analysis. By contrast, the firm’s revenue function in our problem is neither concave in its price nor satisfies any
aforementioned properties, and it even becomes non-smooth under asymmetric reference effects. Besides, due to the dynamic nature of reference price, the standard concept of equilibrium fails to characterize the convergence. To address the issues, we adopt the concept of SNE to jointly capture the market equilibrium and stability. For both loss-neutral and loss-averse reference effects, we demonstrate that the proposed algorithms provably converge to the SNE.

3. Problem Formulation

3.1. MNL Demand Model with Reference Effects

We study a duopoly price competition that involves potentially asymmetric reference price effects, where two firms each offer a substitutable product, labeled as $H$ and $L$, respectively. Both firms set prices concurrently at each period throughout an infinite time horizon. To capture the cross-product effects between two firms, we employ a multinomial logit (MNL) model. The consumers’ utility at period $t$, which depends on the posted price $p^t_i$ and reference price $r^t_i$, is defined as

$$U_i(p^t_i, r^t_i) = u_i(p^t_i, r^t_i) + \epsilon^t_i = a_i - b_i \cdot p^t_i + c^+_i \cdot (r^t_i - p^t_i)_+ + c^-_i \cdot (r^t_i - p^t_i)_- + \epsilon^t_i, \quad \forall i \in \{H, L\},$$

where $u_i(p^t_i, r^t_i)$ is the deterministic component with given parameters $(a_i, b_i, c^+_i, c^-_i)$, and $\epsilon^t_i$ denotes the random fluctuation following the i.i.d. standard Gumbel distribution. The notations $(\cdot)_+ := \max\{\cdot, 0\}$ and $(\cdot)_- := \min\{\cdot, 0\}$ are adopted to account for consumers’ potentially asymmetric reactions to discounts and surcharges, respectively. According to the random utility maximization theory (McFadden 1974), the market share at period $t$ with the posted price $p^t = (p^t_H, p^t_L)$ and reference price $r^t = (r^t_H, r^t_L)$ is given by

$$d_i(p^t, r^t) = d_i((p^t_H, p^t_L), (r^t_H, r^t_L)) = \frac{\exp u_i(p^t_i, r^t_i)}{1 + \exp(u_i(p^t_i, r^t_i) + \exp(u_{-i}(p^t_{-i}, r^t_{-i}))}, \quad \forall i \in \{H, L\},$$

where the subscript “$-i$” denotes the other product besides product $i$. Consequently, the expected revenue for each firm at period $t$ can be expressed as

$$\Pi_i(p^t, r^t) = \Pi_i((p^t_H, p^t_L), (r^t_H, r^t_L)) = p^t_i \cdot d_i(p^t, r^t), \quad \forall i \in \{H, L\}.$$ 

We provide the interpretation for parameters $(a_i, b_i, c^+_i, c^-_i)$ in Eq. (1). For product $i \in \{H, L\}$, $a_i$ signifies the intrinsic value of product $i$, and $b_i$ quantifies the consumers’ price sensitivity. In the presence of asymmetric reference effects, $c^+_i$ and $c^-_i$ correspond to the reference price sensitivities to discounts and surcharges, respectively. Specifically, when the offered price is lower than the internal reference price ($r^t_i > p^t_i$), consumers perceive it as discounts or gains, whereas the price above the reference price ($r^t_i < p^t_i$) is regarded as surcharges or losses. Acknowledging the potential asymmetry in consumer responses to gains and losses, we categorize reference effects into three types: loss-averse, loss-neutral, and gain-seeking, as detailed in Definition 1.
Definition 1 (Product-specific Reference Effects). For each product $i \in \{H, L\}$, consumers are loss-averse when $c_i^+ < c_i^-$ (more responsive to losses), loss-neutral when $c_i^+ = c_i^-$ (equally responsive to losses and gains), or gain-seeking when $c_i^+ > c_i^-$ (more responsive to gains).

In Definition 1, we classify the reference effect for each product based on its relative sensitivities to gains and losses, a categorization method widely employed in the related literature (see, e.g., Kopalle et al. (1996), Köbberling and Wakker (2005), Popsesu and Wu (2007), Hu et al. (2016), Chen et al. (2017), Chen and Nasiry (2020), Jiang et al. (2022)).

We assume the sensitivities to be positive, i.e., $b_i, c_i^+, c_i^- > 0$ for $i \in \{H, L\}$, which agrees with consumer behaviors toward substitutable products and is a common practice in similar pricing problems (see, e.g., Hu et al. (2016), Chen et al. (2017), Golrezai et al. (2020), Guo et al. (2022), Guo and Shen (2023)). Precisely, this assumption guarantees that a rise in $p_t^i$ leads to a decrease in $u_i(p_t^i, r_t^i)$, thereby reducing its own market share $d_i(p_t^i, r_t^i)$ and augmenting that of the competing product $d_{-i}(p_t^i, r_t^i)$. Conversely, an increase in the reference price $r_t^i$ boosts the utility for product $i$, subsequently affecting consumer demands in the inverse manner. Moreover, we stipulate the feasible range for both price and reference price to be $\mathcal{P} = [p, \overline{p}]$, where $p, \overline{p} > 0$ represent the lower and upper bounds of the price, respectively. This price constraint is well-aligned with the real-world instances of price floors and ceilings. Its validity is further justified by the widespread application in the literature on price optimization with reference effects (see, e.g., Hu et al. (2016), Chen et al. (2017), Golrezai et al. (2020)).

We formulate the reference price using the brand-specific past prices (PASTBRSP) model proposed by Briesch et al. (1997), which posits that the reference price is product-specific and memory-based. Compared to other reference price models evaluated in the same study, this model is preferred for its superior performance in terms of fit and prediction. In particular, the reference price for product $i$ is constructed by applying exponential smoothing to its own historical prices, where the memory parameter $\alpha \in [0, 1]$ controls the rate at which the reference price evolves. Starting with the initial price and reference price set at $(p_0^i, r_0^i)$ for product $i$, the reference price update at period $t$ can be described as

$$r_{t+1}^i = \alpha \cdot r_t^i + (1 - \alpha) \cdot p_t^i, \quad \forall i \in \{H, L\}, \quad t \geq 0. \quad (4)$$

The exponential smoothing technique is among the most prevalent and empirically validated reference price update mechanisms in the existing literature (see, e.g., Greenleaf (1995), Kopalle et al. (1996), Popsesu and Wu (2007), Nasiry and Popsesu (2011), Hu et al. (2016), Chen et al. (2017)). We remark that the theories established in this work are readily generalizable to the scenario with time-varying memory parameters $\alpha$, and here we present with the static $\alpha$ for the sake of clarity.
3.2. Opaque Market Setup

The opaque setting is specified as follows: each firm operates without information regarding its competitors, its own reference prices, and the scheme behind reference price updates. That is to say, each firm \( i \) only maintains knowledge of its previously posted price and its own sensitivity parameters \( (b_i, c_i^+, c_i^-) \). Furthermore, we assume that firms know the overall market size (i.e., the largest potential demand including non-purchase quantity). This information does not disclose the firm’s market share to its competitor, thereby not compromising the goal of preserving the firm’s confidentiality. Under this configuration, each firm \( i \) cannot directly compute either \( d_i(p^i, r^i) \) or \( d_{-i}(p^i, r^i) \) from the expression in Eq. (2). Yet, it is legitimate to assume that at the end of each period \( t \), firm \( i \) can access its realized sales volume. Together with the knowledge of the overall market size, this allows the firm to deduce its market share \( d_i(p^i, r^i) \). The realized demand can be viewed as the market feedback in response to the prices set by both firms at the beginning of the period. We underline that in this non-transparent and non-cooperative market, such feedback mechanisms are pivotal for price optimization. In Sections 4 and 5, we show that as a consequence of such an opaque setup, the firms are able to adjust prices by leveraging the first-order information.

Additionally, we comment that it is feasible to estimate the parameters and overall market size from historical data using existing methods in the literature. With uncensored data, where both purchase and non-purchase data are available, these estimations are straightforward. Such cases typically arise in online shopping, where the non-purchase data is tracked by platforms. When comes to the censored data, where the non-purchase quantities are unavailable, parameter calibration and market size approximation require more sophisticated approaches, e.g., the generalized expectation-maximization gradient method proposed by Wang and Wang (2017). The estimation and the impacts of its resulting errors are further elaborated as an extension of this work in Section 6.

3.3. Market Equilibrium and Stability

The goal of our paper is to devise a simple and intuitive pricing update mechanism that ensures the firms achieve the equilibrium and market stability in the long run while protecting their privacy. Before introducing the notion of stable equilibrium considered in this work, we first define the equilibrium pricing policy denoted by \( p^*(r) = (p^*_H(r), p^*_L(r)) \), which is a function that maps reference price to price and achieves the pure strategy Nash equilibrium in the single-period game. Mathematically, the equilibrium pricing policy satisfies that

\[
p^*_i(r) = \arg \max_{p_i \in \mathcal{P}} p_i \cdot d_i((p_i, p^*_{-i}(r)), r), \quad \forall i \in \{H, L\}.
\]

(5)

Next, we formally introduce the concept of stationary Nash equilibrium, which is utilized to jointly characterize the market equilibrium and stability. Note that a similar notion has also been studied by Golrezaei et al. (2020) and Guo and Shen (2023).
Definition 2 (Stationary Nash equilibrium). A price vector $p^\star\star$ is considered a stationary Nash equilibrium (SNE) if $p^\star(p^\star\star) = p^\star\star$, i.e., the equilibrium price is equal to its reference price.

From Eq. (5) and Definition 2, we observe that an SNE possesses the following two properties:

- **Equilibrium.** The revenue function for each firm $i \in \{H, L\}$ satisfies $\Pi_i (p_i, p^\star\star - i) \leq \Pi_i (p^\star\star, p^\star\star)$, i.e., when their reference prices and the price of the other firm are equal to the SNE price, the best-response price for firm $i$ is its SNE price $p^\star\star_i$.

- **Stability.** If the price and the reference price attain the SNE at some period $t$, they would remain unchanged in the following periods, i.e., $p^t = r^t = p^\star\star$ implies that $r^\tau = p^\tau = p^\star\star$, $\forall \tau \geq t + 1$.

As a result, when the market reaches the SNE, the firms have no incentive to deviate, and the market remains stable in subsequent competitions. The next proposition establishes the properties of SNE(s) across all types of reference effects, with its proof detailed in Appendix C.1.

**Proposition 1.** Define $S$ as the set of SNE(s). Then, the following statements hold:

- In the presence of any product with gain-seeking reference effects, $S$ is empty.
- In the loss-neutral scenario, $S$ is a singleton, i.e., there exists a unique SNE $p^\star\star = (p^\star\star_H, p^\star\star_L)$.
- In the loss-averse scenario, there exist multiple SNEs and $S$ can be non-convex.

Furthermore, when $S \neq \emptyset$, any SNE $p \in S$ can be bounded as

$$\frac{1}{b_i + c_i^-} < p_i < \frac{1}{b_i + c_i^+} + \frac{1}{b_i} W\left(\frac{b_i}{b_i + c_i^+} \exp\left(a_i - \frac{b_i}{b_i + c_i^+}\right)\right), \quad \forall i \in \{H, L\}, \quad (6)$$

where $W(\cdot)$ is the Lambert $W$ function (see definition in Eq. (C.9)).

As indicated by Proposition 1, the presence of one or more gain-seeking products implies the non-existence of SNE, suggesting that the long-run market equilibrium and stability cannot be achieved simultaneously. Therefore, the remaining part of the paper will be devoted to the loss-neutral and loss-averse scenarios. To avoid ambiguity in the types of scenarios, we provide precise definitions: a loss-neutral scenario occurs when both products have loss-neutral reference effects; a loss-averse scenario arises when at least one product displays loss-averse reference effects, with the other being either loss-neutral or loss-averse. Consequently, the three cases listed in Proposition 2 are mutually exclusive and cover all possible combinations of reference effect types.

To complement Proposition 1, we illustrate the non-convexity of set $S$ when products with loss-averse reference effects are present, as shown in Figure 1 (the loss-aversion is manifested by the inequalities $c_i^+ < c_i^-$, $\forall i \in \{H, L\}$, as outlined in the figure caption). The grey region in Figure 1a depicts the set of SNEs, demonstrating its non-convex nature to some extent. The four colored curves correspond to the four equations that define the set $S$ (see Eqs. (C.3) and (C.11)). A closer inspection of the area surrounding the upper green curve, as displayed in Figure 1b, further verifies this non-convexity, with the black dashed line being a straight path between two vertices.
Figure 1 Illustration for the Non-convexity of $S$ in the Loss-averse Scenario. The Shaded Region Represents $S$ and the Colored Curves Represent the Boundaries of $S$.

(Parameters: $(a_H, b_H, c_H^+, c_H^-) = (5.88, 4.20, 1.17, 3.63)$ and $(a_L, b_L, c_L^+, c_L^-) = (5.32, 1.16, 1.77, 4.12)$)

Without loss of generality, we assume that the feasible price range $\mathcal{P}^2 = [p_L, p_U]^2$ is sufficiently large to contain the set of SNE(s), i.e., $S \subseteq [p_L, p_U]^2$. Proposition 1 provides a quantitative characterization for this assumption: it suffices to choose the price lower bound $p_L$ to be any real number between $0, \min_{i \in \{H,L\}} \{1/(b_i + c_i^-)\}$, and the price upper bound $p_U$ can be any value such that

$$p_U \geq \max_{i \in \{H,L\}} \left\{ \frac{1}{b_i + c_i^+} + \frac{1}{b_i} W \left( \frac{b_i}{b_i + c_i^+} \exp \left( a_i - \frac{b_i}{b_i + c_i^+} \right) \right) \right\}. \quad (7)$$

This assumption is mild since the bound in Eq. (7) is independent of both the price and reference price, and it does not grow exponentially fast with respect to any parameters. Hence, there is no need for $p_U$ to be excessively large. For example, when $a_H = a_L = 10$ and $b_H = b_L = c_H^+ = c_L^+ = 1$, the upper bound from Eq. (7) becomes $p_U \geq 7.3785$. Further details on the structure and computation of the equilibrium pricing policy and the SNE(s) can be found in Appendices B and C.

4. **Loss-neutral: Online Projected Gradient Ascent**

In this section, we delve into the algorithm design for the loss-neutral scenario, with the objective of driving the convergence of both price and reference price to the SNE. Because of the symmetric reference effect for loss-neutral consumers, we simplify the notation of the reference price sensitivity and denote $c_i := c_i^+ = c_i^-$, $\forall i \in \{H, L\}$. 
As studied in Guo and Shen (2023), with perfect information, firms are able to compute both the equilibrium pricing policy (see definition in Eq.(5)) and the best-response pricing policy in each period, where the best-response policy refers to the firm setting its current price as the optimal response to the competitor’s previous-period price and both firms’ current reference prices. They demonstrate that the market dynamics under these pricing policies would stabilize in the long run. However, with partial information, firms are incapable of computing either the equilibrium or the best-response policy, as they lack the essential information about their competitors. Thus, in an opaque environment, a practical yet effective strategy for firms to boost revenues and move toward equilibrium is to dynamically modify prices based on the market feedback.

In light of the success of gradient-based algorithms in the online learning literature (see, e.g., Zinkevich (2003), Hazan et al. (2007), Abernethy et al. (2011)), we propose the Online Projected Gradient Ascent (OPGA) method, as outlined in Algorithm 1. Specifically, at each period, both firms update their current prices using the first-order derivatives of their log-revenues with the same learning rate (see Lines 3–5). Importantly, the firms are not required to know their own reference price when executing the OPGA algorithm, and the reference price update in Line 6 is automatically handled by the market.

\textbf{Algorithm 1:} Online Projected Gradient Ascent (OPGA)

1. \textbf{Input:} Initial reference price $r^0 = (r^0_H, r^0_L)$, initial price $p^0 = (p^0_H, p^0_L)$, and step-sizes $\{\eta_t\}_{t \geq 0}$.
2. for $t = 0, 1, 2, \ldots$ do
3.  \hspace{1em} for $i \in \{H, L\}$ do
4.  \hspace{2em} Compute derivative from price and market share of firm $i$ at period $t$:
5.  \hspace{2em} \begin{align}
6.  D^t_i &\leftarrow \frac{\partial \log(\Pi_i(p^t, r^t))}{\partial p_i} = \frac{1}{p^t_i} + (b_i + c_i) \cdot d_i(p^t, r^t) - (b_i + c_i). \tag{8}
7.  \end{align}
8.  \hspace{2em} Update posted price: $p^t_{i+1} \leftarrow \text{Proj}_p (p^t_i + \eta^t \cdot D^t_i)$.
9.  \hspace{1em} end
10. Update reference price: $r^{t+1} \leftarrow \alpha \cdot r^t + (1 - \alpha) \cdot p^t$.
11. end

It is noteworthy that the difference between the derivative of the log-revenue in Eq. (8) and that of the standard revenue is a scaling factor equal to the revenue itself, i.e.,

\[ \frac{\partial \log(\Pi_i(p, r))}{\partial p_i} = \frac{1}{\Pi_i(p, r)} \cdot \frac{\partial (\Pi_i(p, r))}{\partial p_i}, \quad \forall i \in \{H, L\}. \tag{9} \]

Therefore, the price update in Algorithm 1 can be equivalently viewed as an adaptively regularized gradient ascent using the standard revenue function, where the regularizer at period $t$ is
Moreover, we highlight that by leveraging the structure of the MNL model, each firm $i$ can obtain the derivative of its log-revenue $D_t^i$ in Eq. (8) through the market feedback mechanism, i.e., the most recent demand. In fact, computing the derivative $D_t^i$ is identical to querying the first-order oracle, which is a common assumption in the optimization literature (Nesterov 2003). Furthermore, even without access to the historical prices and the market feedback, the derivative in Eq. (8) can also be acquired through a minor perturbation of the posted price, which makes the algorithm applicable in a variety of scenarios.

Our problem is related to two streams of existing research: multi-player games and discrete nonlinear systems. Despite of the similarity, established theories and techniques in these areas are insufficient for the convergence analysis of Algorithm 1. Below, we briefly introduce how our problem can be transformed into these formulations and discuss the underlying challenges in the analysis. For a more comprehensive explanation, we direct readers to Appendix A.

- **Four-player Game with No Underlying Dynamic State.** By adding two virtual firms to represent the reference prices, the game can be converted into a standard four-player online game, effectively eliminating the underlying dynamic state, i.e., reference price. However, analyzing this four-player game presents a great challenge due to the absence of certain favorable properties in the objective functions of the real firms (i.e., revenues). These properties, such as strong monotonicity (Lin et al. 2020) or variational stability (Mertikopoulos and Zhou 2019), are typically instrumental in proving the convergence of online games. Moreover, while the real firms have the flexibility to dynamically adjust their step-sizes, the learning rate for virtual firms is fixed to the constant $(1 - \alpha)$, where $\alpha$ is the memory parameter for reference price updates. This disparity also hinders the direct application of the existing results from multi-agent online learning literature, as their results typically require the step-sizes of all agents to either diminish at comparable rates or remain as a sufficiently small constant (see, e.g., Nagurney and Zhang (1995), Scutari et al. (2010), Bravo et al. (2018), Mertikopoulos and Zhou (2019)).

- **Discrete Nonlinear System.** The second approach involves translating the OPGA algorithm into a discrete nonlinear system by treating $(p^{t+1}, r^{t+1})$ as a vector-valued function of $(p^t, r^t)$, i.e., $(p^{t+1}, r^{t+1}) = f(p^t, r^t)$ for some function $f(\cdot)$. In this context, analyzing the convergence of Algorithm 1 is equivalent to examining the stability of the fixed point of $f(\cdot)$, which is related to the spectral radius of the Jacobian matrix $\nabla f(p^*, p^*)$ (Arrowsmith et al. 1990, Olver 2015). However, the SNE lacks a closed-form expression, making it difficult to calculate the eigenvalues of $\nabla f(p^*, p^*)$. In addition, the function $f(\cdot)$ is non-smooth due to the presence of the projection operator and can become non-stationary when the firms adopt time-varying step-sizes, such as diminishing step-sizes. Moreover, typical results in dynamical systems only guarantee local convergence (Khalil 2002), i.e., the asymptotic stability of the fixed point, whereas our goal is to establish the global convergence of both the price and reference price.
In the following section, we show that the OPGA algorithm with diminishing step-sizes converges
to the unique SNE by exploiting the distinctive properties of our model. This convergence result
indicates that the OPGA algorithm provably achieves no-regret learning, i.e., in the long run, the
algorithm performs at least as well as the best fixed action in hindsight. However, we remark that
the reverse direction is not necessarily true: being no-regret does not guarantee the convergence
at all, let alone the convergence to an equilibrium (see, e.g., Mertikopoulos et al. (2018, 2019)). In
fact, the agents may exhibit entirely unpredictable and chaotic behaviors under a no-regret policy
(Palaiopanos et al. 2017), with the only exception being the finite game, where players compete
for finitely many rounds (note that our problem does not fall under this category).

4.1. Convergence Result
In this section, we investigate the convergence properties of the OPGA algorithm. We first establish
the global last-iterate convergence of Algorithm 1 to the unique SNE in Theorem 1. Subsequently,
we show in Theorem 2 that this convergence exhibits a rate of $O(1/t)$, provided that the step-sizes
are selected appropriately. Lastly, we prove in Theorem 3 that Algorithm 1 achieves an $O(\sqrt{T})$
dynamic regret. The formal proofs for this section can be found in Appendices D, E, and F, with
the auxiliary lemmas located in Appendix I.

**Theorem 1 (Global Convergence of OPGA).** In the loss-neutral scenario, let the step-
sizes $\{\eta^t\}_{t \geq 0}$ be a non-increasing sequence such that $\lim_{t \to \infty} \eta^t = 0$ and $\sum_{t=0}^{\infty} \eta^t = \infty$ hold. Then,
the price paths and reference price paths generated by Algorithm 1 with step-sizes $\{\eta^t\}_{t \geq 0}$ converge
to the unique stationary Nash equilibrium.

Theorem 1 demonstrates the global last-iterate convergence of the OPGA algorithm to the unique
SNE, thus ensuring the market equilibrium and stability in the long term. Compared to Guo and
Shen (2023) who establish the convergence to SNE under the perfect information setting, Theorem
1 ensures that such convergence can also be achieved in an opaque market where the firms are
reluctant to share the information with their competitors. The only prerequisite for the conver-
geence is the diminishing step-sizes that satisfy $\lim_{t \to \infty} \eta^t = 0$ and $\sum_{t=0}^{\infty} \eta^t = \infty$. This assumption
is widely adopted in the studies of online games (see, e.g., Mertikopoulos and Zhou (2019), Bravo
et al. (2018), Lin et al. (2021)). Since the firms are likely to become more acquainted with their
competitors through repeated interactions, it is reasonable for them to be more conservative in
price adjustments over time, leading to a gradual reduction in learning rates.

As discussed in Sections 2 and 4, the existing methods in the online game and dynamical system
literature are not applicable to our problem due to the absence of standard structural properties
and the presence of the underlying dynamic state, i.e., reference price. Consequently, we develop a
novel analysis to prove Theorem 1, leveraging the characteristic properties of the MNL model. We provide a proof sketch below, with the complete proof deferred to Appendix D.

**High-level Proof Overview.** In a nutshell, our convergence analysis of the OPGA algorithm primarily comprises the following two parts:

- **Part 1** (Appendix D.1). We show that the price path \( \{ p_t \}_{t \geq 0} \) would enter the neighborhood \( N_1^\epsilon := \{ p \in P^2 \mid \varepsilon(p) < \epsilon \} \) infinitely many times for any \( \epsilon > 0 \), where \( \varepsilon(\cdot) \) is a weighted \( \ell_1 \)-distance function defined as \( \varepsilon(p) := \frac{|p_H^* - p_H|}{(b_H + c_H)} + \frac{|p_L^* - p_L|}{(b_L + c_L)} \). To prove Part 1, we divide \( P^2 \) into four quadrants with \( p^* \) as the origin. Then, employing a contradiction-based argument, we suppose that the price path only visits \( N_1^\epsilon \) finitely many times. Yet, we show that such a premise forces the price path to oscillate between adjacent quadrants and ultimately converge to the SNE, which violates the initial assumption.

- **Part 2** (Appendix D.2). We show that when the price path \( \{ p_t \}_{t \geq 0} \) enters the \( \ell_2 \)-neighborhood \( N_2^\epsilon := \{ p \in P^2 \mid \| p - p^{**} \|_2 < \epsilon \} \) for some sufficiently small \( \epsilon > 0 \) and with small enough step-sizes, the price path will remain in \( N_2^\epsilon \) in subsequent periods. The proof of Part 2 relies on a local property of the MNL demand around the SNE, with which we demonstrate that the price update provides adequate ascent to ensure the price path stays within the target neighborhood.

Owing to the equivalence between different distance functions in Euclidean spaces (Rudin 1987), there exists a constant ratio \( \rho > 0 \) such that \( N_1^{\epsilon/\rho} \subset N_2^\epsilon \) for every \( \epsilon > 0 \). Hence, as we can choose arbitrarily small \( \epsilon \) in Part 1, the above two parts jointly imply the convergence of the price paths to the SNE. Since the reference price results from the exponential smoothing of historical prices, its convergence follows from that of the price path.

It is worth noting that Golrezaei et al. (2020, Theorem 5.1) also employs a two-part proof and shows the asymptotic convergence of online mirror descent (Zhou et al. 2017, Hoi et al. 2021) to the SNE in the linear demand setting. However, their analysis heavily depends on the linear structure of the demand, which ensures that a property Golrezaei et al. (2020, Lemma 9.1) similar to the variational stability is globally satisfied. In contrast, such properties no longer hold for the MNL model considered in this paper. Therefore, we come up with two distinct distance functions for the two parts of the proof, respectively. Moreover, the proof in Golrezaei et al. (2020) relies on an assumption concerning the relation between the sensitivity parameters, whereas our convergence analysis only necessitates the most basic assumption: the positivity of the sensitivity parameters. Without this, the model becomes counter-intuitive for substitutable products.

**Theorem 2 (Global Convergence Rate).** In the loss-neutral scenario, there exists constants \( d_\eta, d_p, d_r > 0 \) such that when both firms adopt Algorithm 1 with step-sizes \( \eta^t = d_\eta/t \) for all \( t \geq 1 \), the convergence rate of \( \{ (p^t, r^t) \}_{t \geq 0} \) is described as

\[
\| p^{**} - p^t \|_2^2 \leq \frac{d_p}{t}, \quad \| p^{**} - r^t \|_2^2 \leq \frac{d_r}{t}, \quad \forall t \geq 1.
\]
Theorem 2 improves the result of Theorem 1 by demonstrating the non-asymptotic convergence rate for the price and reference price. Recall that in our problem, the MNL demand model leads to non-concave revenue functions for both firms. Although such non-concavity generally anticipates a slower convergence, our rate of $O(1/t)$ matches that in Golrezaei et al. (2020, Theorem 5.2), where their problem exhibits concavity due to the linear demand.

The proof of Theorem 2 is primarily based on the local convergence rate when the price vector remains in a specific neighborhood $N^2_{\epsilon_0}$ of the SNE after some period $T_{\epsilon_0}$. As the choice $\eta^t = \Theta(1/t)$ ensures that $\lim_{t \to \infty} \eta^t = 0$ and $\sum_{i=0}^{\infty} \eta^t = \infty$, the existence of such a period $T_{\epsilon_0}$ is guaranteed by Theorem 1. Utilizing an inductive argument, we first show that the difference between the price and the reference price decreases at a faster rate of $O(1/t^2)$, i.e., $r^t - p^t = O(1/t^2)$, $\forall t \geq 1$ (see Eq. (E.11)). Then, by exploiting a local property around the SNE (see Lemma EC.4), we use another induction to establish the convergence rate of the price path after it enters the neighborhood $N^2_{\epsilon_0}$. In the meanwhile, the convergence rate of the reference price path can be determined through a triangular inequality, i.e., $\|p^{**} - r^t\|_2 = \|p^{**} - p^t + p^t - r^t\|_2 \leq 2\|p^{**} - p^t\|_2 + 2\|p^t - r^t\|_2$. Finally, due to the boundedness of the feasible price range $P^2$, we can obtain the global convergence rate for all $t \geq 1$ by choosing sufficiently large constants $d_p$ and $d_r$. The complete proof of Theorem 2 is in Appendix E.

**Remark 1 (Constant step-sizes).** The convergence analysis presented in this work can be readily extended to accommodate constant step-sizes. Particularly, our theoretical framework supports the selection of $\eta^t = O(\epsilon_0)$, where $\epsilon_0$ refers to the size of the neighborhood used in Theorem 2. When $\eta^t = O(\epsilon_0)$, an approach akin to Part 1 can be employed to demonstrate that the price path enters the neighborhood $N^1_{\epsilon}$ infinitely many times for any $\epsilon = O(\epsilon_0)$. Subsequently, by exploiting the local property in a manner similar to Part 2, we can establish that once the price path enters the neighborhood $N^2_{\epsilon}$ for some $\epsilon = O(\epsilon_0)$, it will remain within that neighborhood.

We conduct numerical experiments to further illustrate the convergence behavior of OPGA. The pair of examples (refer to Figures 2a, 2b, and 2c), which differ solely in the choices of step-sizes, highlight the crucial role of step-sizes in attaining convergence. In particular, Examples 1 and 2 (see Figures 2a and 2b) affirms Theorem 1 by demonstrating that the price and reference price paths converge to the unique SNE when the chosen diminishing step-sizes fulfill the criteria specified by Theorem 1. By contrast, the over-large constant step-sizes employed in Example 3 (see Figure 2c) fail to achieve convergence and lead to cyclic patterns in the long run. In Example 4 (see Figure 2d), we compare the OPGA algorithm and the repeated implementation of the equilibrium pricing policy (see definition in Eq. (5)) by plotting the reference price trajectories produced by both methods. Figure 2d conveys that the two algorithms reach the SNE at a comparable rate,
Figure 2  Price and Reference Price Paths for Examples 1, 2, 3, and 4.

(Parameters: \((a_H, b_H, c_H) = (8.70, 2.00, 0.82), (a_L, b_L, c_L) = (4.30, 1.20, 0.32), (r_0^H, r_0^L) = (0.10, 2.95), (p_0^H, p_0^L) = (4.85, 4.86), \) and \(\alpha = 0.90.\))

(a) Example 1: Convergence with \(\eta^i = 3/t.\)  
(b) Example 2: Convergence with \(\eta^i = 1/\sqrt{t}.\)  
(c) Example 3: Cyclic Pattern with \(\eta^i = 1.\)  
(d) Example 4: OPGA vs. Equilibrium Policy.

which indicates that the OPGA algorithm allows the firms to attain the equilibrium and stability as though they operate under a transparent market while preventing information exposure.

Next, we derive the regret of Algorithm 1. The standard (static) regret compares an online algorithm with the best fixed decision in hindsight. The rationale behind the static metric is that this best fixed decision performs adequately well across all iterations. Nevertheless, given the evolving underlying state (i.e., reference price) in our context, such an assumption becomes overly optimistic. To address this limitation, we choose the dynamic regret as the performance measure, defined as the difference between cumulative revenue attained by an online algorithm and a sequence of best decisions in hindsight. The dynamic regret is strictly stronger than the static
regret in a non-stationary environment, as the latter only benchmarks against the single best fixed action over all rounds. Let $\text{D-Regret}_i(T)$ denote the dynamic regret for firm $i$ over $T$ periods, which is defined as

$$\text{D-Regret}_i(T) := \sum_{t=1}^{T} \max_{p_i \in P} \left\{ \Pi_i((p_i, p_{-i}^t), r^t) - \Pi_i(p^t, r^t) \right\}.$$  \hspace{1cm} (11)

**Theorem 3 (Dynamic Regret Bound).** In the loss-neutral scenario, when both firms adopt Algorithm 1 with step-sizes $\{\eta^t\}_{t \geq 0}$ specified by Theorem 1, the dynamic regret of each firm $i \in \{H, L\}$ grows in a sublinear rate, i.e.,

$$\lim_{T \to \infty} \frac{1}{T} \times \text{D-Regret}_i(T) = 0.$$ \hspace{1cm} (12)

Furthermore, when the step-sizes are specified by Theorem 2, it holds that

$$\text{D-Regret}_i(T) = 2 \left[ \left( p(b_i + c_{-i}) + \ell_{p,i} \right) \sqrt{d_p} + \left( p(c_i + \ell_{r,i}) \sqrt{d_r} \right) \right. \sqrt{T} = O(\sqrt{T}),$$ \hspace{1cm} (13)

where $\ell_{p,i} := \sqrt{(1 + (b_i + c_i)p/4)^2 + ((b_i + c_{-i})p/4)^2}$, $\ell_{r,i} := \sqrt{(c_i p/4)^2 + (c_{-i} p/4)^2}$, and $d_p, d_r$ are the constants introduced in Theorem 2.

Theorem 3 shows that Algorithm 1 yields a sublinear dynamic regret for any non-increasing step-sizes satisfying $\lim_{t \to \infty} \eta^t = 0$ and $\sum_{t=0}^{\infty} \eta^t = \infty$. Moreover, the dynamic regret grows in the rate of $O(\sqrt{T})$ when the step-sizes are selected as $\eta^t = d_\eta/t$ for all $t \geq 1$. We note that the proof of Theorem 3 is built upon the last-iterate convergence result in Theorems 1 and 2. Thus, it further confirms that the last-iterate convergence is stronger than purely being no-regret. The complete proof of Theorem 3 can be found in Appendix F.

### 5. Loss-averse: Conservative Online Projected Gradient Ascent

In transitioning from symmetric to asymmetric reference effects, the non-existence of SNE associated with gain-seeking effects (see Proposition 1) directs our attention to the loss-averse scenario, where consumers exhibit the loss-aversion toward at least one product, i.e., $c_{-i} \geq c_{i}^+$ for $i \in \{H, L\}$, with at least one inequality being strict. According to Proposition 1, with loss-averse reference effects, the set of SNEs $S$ is not confined to a single point and could possibly be non-convex. Thus, while it is natural to assess the convergence using the distance between the current price and the unique SNE price $p^{**}$ in the loss-neutral case, such distance-based metrics are no longer well-defined for loss-averse scenarios due to the non-convexity of $S$. Inspired by the criteria for stationarity found in the non-convex non-smooth optimization literature (see, e.g., Zhang et al. (2020), Li et al. (2020)), we propose a novel metric to evaluate the convergence under asymmetric
reference effects, taking both price and reference price into account. For any given pair of \((p, r)\), we define the metric \(E(p, r)\) as follows:

\[
E(p, r) := \|p - r\|_2 + \sum_{i \in \{H, L\}} \text{dist} \left( 0, \text{Hull} \left\{ D_i^-(p, r), D_i^+(p, r) \right\} \right), \tag{14}
\]

where \(\text{dist}(\cdot, \cdot)\) denotes the Euclidean distance function, \(\text{Hull}(\cdot, \cdot)\) refers to the convex hull of the input points, and the functions \(D_i^+(\cdot, \cdot)\) and \(D_i^-(\cdot, \cdot)\) are defined as

\[
D_i^+(p, r) := \frac{1}{p_i} + (b_i + c_i^+) \cdot d_i(p, r) - (b_i + c_i^+) \tag{15a}
\]

\[
D_i^-(p, r) := \frac{1}{p_i} + (b_i + c_i^-) \cdot d_i(p, r) - (b_i + c_i^-). \tag{15b}
\]

In the loss-averse scenario, given that \(c_i^- \geq c_i^+\), the term \(\text{dist}(0, \text{Hull}\{D_i^-(p, r), D_i^+(p, r)\})\) actually measures the distance between 0 and the interval \([D_i^-(p, r), D_i^+(p, r)]\).

The metric \(E(p, r)\) consists of two components, where the first one quantifies the distance between the current price and the reference price, and the second part serves as an indicator for a stationary point in the non-smooth revenue function. For any specified pair of \((p, r)\) with \(p_i \neq r_i\), it is evident from Eq. (15) that exactly one of \(D_i^+(p, r)\) and \(D_i^-(p, r)\) is equal to the derivative of firm \(i\)'s log-revenue. Meanwhile, the other quantity can be regarded as the virtual derivative assuming the opposite reference price sensitivity (with \(d_i(p, r)\) still being the actual market share). For instance, when \(p_i < r_i\), the effective reference price sensitivity should be \(c_i^+\), and thereby \(D_i^-(p, r) = \partial \log (\Pi_i(p, r)) / \partial p_i\) is the true derivative. The remaining \(D_i^+(p, r)\) is the virtual derivative and can be obtained by differentiating \(\log (\Pi_i(p, r))\) as if the effective reference price sensitivity is \(c_i^-\). We remark that when \(p_i = r_i\), the terms \(D_i^+(p, r)\) and \(D_i^-(p, r)\) correspond to the left-hand and right-hand derivatives of the log-revenue, respectively. In the next proposition, we confirm that \(E(p, r)\) is a well-defined metric for determining the convergence to SNEs. The proof of the proposition can be found in Appendix C.2.

**Proposition 2.** The set of SNEs can be equivalently expressed as \(S = \{p \in P^2 \mid E(p, p) = 0\}\).

Thus, given the compactness of \(P^2\) and the continuity of \(E(\cdot, \cdot)\), Proposition 2 ensures that a price vector \(p\) is sufficiently close to \(S\) if and only if \(E(p, p)\) is small enough. Motivated by this implication, we introduce the concept of \(\epsilon\)-approximated SNE, abbreviated as \(\epsilon\)-SNE.

**Definition 3 (\(\epsilon\)-SNE).** For any \(\epsilon > 0\), a price vector \(p\) is an \(\epsilon\)-SNE if \(E(p, p) < \epsilon\).

Below, we focus on constructing an algorithm to find an \(\epsilon\)-SNE for any given \(\epsilon > 0\). Given Proposition 2, when \(\epsilon\) is sufficiently small, the identified point is guaranteed to be a good approximation for SNEs. To deal with the non-smoothness induced by the asymmetric reference effects, we propose
**Algorithm 2**: Conservative Online Projected Gradient Ascent (C-OPGA)

1. **Input**: Initial reference price $r^0 = (r^0_H, r^0_L)$, initial price $p^0 = (p^0_H, p^0_L)$, step-sizes $\{\eta^t\}_{t=0}^{\infty}$, and threshold $\epsilon_1$. 

2. for $t = 0, 1, 2, \ldots$ do 

3. 

4. for $i \in \{H, L\}$ do 

5. Compute the true and virtual derivatives from price and market share at period $t$: 

   $$D^i_{t,+} \leftarrow D^i_t(p^t, r^t) = \frac{1}{p^t_i} + (b_i + c^+_i) \cdot d_i(p^t, r^t) - (b_i + c^+_i)$$  

   $$D^i_{t,-} \leftarrow D^i_t(p^t, r^t) = \frac{1}{p^t_i} + (b_i + c^-_i) \cdot d_i(p^t, r^t) - (b_i + c^-_i)$$  

6. if $D^i_{t,+} > -\epsilon_1$ and $D^i_{t,-} < \epsilon_1$ then  

   // Stopping criteria for price updates. 

7. Maintain posted price: $p^{t+1}_i \leftarrow p^t_i$. 

8. else 

9. Compute the convex combination of the true and virtual derivatives at period $t$: 

   $$D^i_t \leftarrow w_i \cdot D^i_{t,+} + (1 - w_i) \cdot D^i_{t,-},$$  

   where the coefficient $w_i$ takes any value within the interval $[0, 1]$. 

10. Update posted price: $p^{t+1}_i \leftarrow \text{Proj}_{P}(p^t_i + \eta^t \cdot D^i_t)$. 

11. end 

12. end 

13. Update reference price: $r^{t+1} \leftarrow \alpha \cdot r^t + (1 - \alpha) \cdot p^t$. 

a modified algorithm called Conservative Online Projected Gradient Ascent (C-OPGA), as detailed in Algorithm 2. The name of the algorithm reflects its more conservative approach compared to the original OPGA, where the C-OPGA integrates a stopping criteria (see Line 5 in Algorithm 2) to refine the process of price updates.

Next, we provide a thorough elucidation of the C-OPGA algorithm. During each period, the firm computes both true and virtual derivatives of its log-revenue (see Eq. (16)). We note that even though firm $i$ may not discern whether $D^i_{t,+}$ or $D^i_{t,-}$ is the true derivative in an opaque environment, it can still compute both quantities through the market feedback $d_i(p^t, r^t)$ and the corresponding sensitivity parameters. These calculated values are then utilized to determine the stopping criteria in Line 5, which can be understood as follows: when the term $\text{dist}(0, \text{Hull}\{D^i_{t,-}, D^i_{t,+}\})$ falls below a predetermined threshold $\epsilon_1$, it triggers a stopping mechanism that pauses the price update for the current round (see Lines 5–6). Otherwise, when the stopping criteria are not met, the firm adjusts
its price according to the convex combination of the true and virtual derivatives, following the procedure outlined in Eq. (17) (see Lines 7–9). Finally, the algorithm concludes with the market performing the reference price update, as specified in Line 10.

**Remark 2.** In Algorithm 2, we assume that the two firms use the same threshold $\epsilon_1$ and time-invariant coefficients $w_H, w_L$. These assumptions are made only for the brevity of the presentation. It is worth noticing that our convergence result in Theorem 4 applies directly to cases when the threshold $\epsilon_1$ is both product-specific and time-variant, and the coefficients $w_H, w_L$ are time-variant.

The subsequent discussion underscores the benefits of introducing the stopping mechanism in C-OPGA. Crucially, the standard vanilla sub-gradient ascent algorithm, which updates the price using the true sub-derivative without any stopping criteria, is not an ascent method even under sufficiently small step-sizes (Boyd et al. 2003). To further complicate the problem, the true derivative in the vanilla sub-gradient algorithm mandates the firms to collect additional information regarding the relative values between the price and reference price. This requirement renders the vanilla approach impractical and unsuitable for our opaque market. The advantage of this stopping mechanism becomes more evident through the two examples presented in Figure 3. We run both the vanilla sub-gradient ascent and the C-OPGA under the same set of parameters and initializations. The price and reference price paths generated by the vanilla sub-gradient ascent (see Figure 3a) initially display signs of convergence, but start to oscillate in a cyclic pattern around the 90th period. This instability of the vanilla algorithm is typically observed when the price and the reference price are in close proximity. Under such conditions, the relative magnitudes of these

**Figure 3**  Comparison of Trajectories between Vanilla Sub-gradient Ascent and C-OPGA
(Parameters: $(a_H, b_H, c_H^+, c_H^-) = (8.19, 1.48, 0.34, 1.50)$, $(a_L, b_L, c_L^+, c_L^-) = (4.59, 1.80, 0.31, 1.14)$, $(r_0^H, r_0^L) = (0.24, 2.19)$, $(p_0^H, p_0^L) = (2.50, 2.75)$, $\alpha = 0.90$, and $\eta^t = 1/\sqrt{t}$.)

(a) Cyclic Pattern with Vanilla Sub-gradient Ascent.  (b) Convergence with C-OPGA.
two values may fluctuate between periods, thereby inducing a jump discontinuity in the derivative due to the discrepancy between $c_i^+$ and $c_i^-$. By contrast, in Figure 3b, the convergence of price and reference price paths under the C-OPGA highlights the effectiveness of the stopping criteria. Indeed, it is easy to verify that the price and reference price paths produced by the C-OPGA in Figure 3b converge to an interior point of the SNE set.

5.1. Convergence Result
As discussed in Section 4, the existing theories fall short of addressing the convergence in the loss-neutral setting. This limitation becomes even more severe in the loss-averse scenario, where the non-smoothness in the revenue function adds another layer of complexity. For instance, this characteristic further exacerbates the challenge of applying discrete nonlinear systems analysis, in which the smoothness is typically required. In spite of these obstacles, we develop an innovative method to show the last-iterate convergence of the C-OPGA algorithm, as formally stated in Theorem 4. Subsequently, we elaborate on the theorem and outline the key ideas behind its proof, with a more comprehensive demonstration reserved for Appendix G.

**Theorem 4 (Global Convergence of C-OPGA).** In the loss-averse scenario, let the step-sizes $\{\eta^t\}_{t \geq 0}$ be a non-increasing sequence such that $\lim_{t \to \infty} \eta^t = 0$ and $\sum_{t=0}^{\infty} \eta^t = \infty$ hold. For any $\epsilon > 0$, let the threshold $\epsilon_1$ satisfy that

$$\left(1 + \max_{i \in \{H,L\}, \sigma \in \{+,-\}} \frac{b_i^\sigma + c_i^\sigma}{b_i - c_i^\sigma}\right) \epsilon_1 \leq \epsilon. \tag{18}$$

Then, the price paths and reference price paths generated by Algorithm 2 with step-sizes $\{\eta^t\}_{t \geq 0}$ and threshold $\epsilon_1$ converge to an $\epsilon$-SNE.

In other words, for any $\epsilon_1 > 0$, Theorem 4 assures the convergence of Algorithm 2 to an $\epsilon$-SNE, with $\epsilon$ adhering to the inequality in Eq. (18). Hence, the firms can attain a highly accurate approximation of SNE by simply setting a sufficiently low threshold $\epsilon_1$ in Algorithm 2. Recall that the firms can indeed use time-variant thresholds $\epsilon_1$ for their stopping criteria (see Remark 2). Therefore, a pragmatic strategy for the firms is to gradually decrease their thresholds along the competition.

**High-level Proof Overview.** To begin with, we divide the feasible price range $\mathcal{P}^2 = [\underline{p}, \overline{p}]^2$ into the following three mutually exclusive and exhaustive regions for each product $i \in \{H,L\}$:

\[
P_{i}^{+1} := \left\{ p \in \mathcal{P}^2 \mid D_i^+(p, p) \geq -\frac{\epsilon_1}{2}, D_i^-(p, p) \leq \frac{\epsilon_1}{2} \right\}, \tag{19}
\]

\[
P_{i}^{+1,+} := \left\{ p \in \mathcal{P}^2 \mid D_i^+(p, p) > \frac{\epsilon_1}{2}, D_i^-(p, p) > \frac{\epsilon_1}{2} \right\}, \tag{20}
\]

\[
P_{i}^{+1,-} := \left\{ p \in \mathcal{P}^2 \mid D_i^+(p, p) < -\frac{\epsilon_1}{2}, D_i^-(p, p) < -\frac{\epsilon_1}{2} \right\}. \tag{21}
\]
Conceptually, this division of \( \mathcal{P}^{i+}_t, \mathcal{P}^{i+}_t, \text{ and } \mathcal{P}^{i-}_t \) relies on the relative position of zero and the interval \( I_i(p) := [D^-_i(p, p), D^+_i(p, p)] \). Specifically, \( \mathcal{P}^{i+}_t \) denotes the instance where the distance between zero and \( I_i(p) \) is no more than \( \epsilon_1/2 \), whereas \( \mathcal{P}^{i+}_t \) (resp. \( \mathcal{P}^{i-}_t \)) indicates that the minimum (resp. maximum) value on the interval \( I_i(p) \) exceeds \( \epsilon_1/2 \) (resp. is less than \(-\epsilon_1/2\)). The latter implies that in \( \mathcal{P}^{i+}_t \) and \( \mathcal{P}^{i-}_t \), the distance from zero to the interval \( I_i(p) \) is greater than \( \epsilon_1/2 \). When considering both products, this division separates the entire feasible price range into nine disjoint and exhaustive regions, implying that an arbitrary price vector must belong to one of the nine regions (i.e., \( p_H \in \mathcal{P}^{i+}_H \cup \mathcal{P}^{i+}_H \cup \mathcal{P}^{i-}_H \) and \( p_L \in \mathcal{P}^{i+}_L \cup \mathcal{P}^{i+}_L \cup \mathcal{P}^{i-}_L \)).

Moving forward in our discussion, we first verify the convergence of the C-OPGA algorithm, followed by confirming that its convergence is indeed toward the set of SNEs. Our analysis anchors on the three mutually exclusive and exhaustive cases, which are classified based on the regions that \( p^t \) belongs to (for some sufficiently large period \( t \)):

- **Case 1.** If \( p^t \in \mathcal{P}^{i+}_H \cap \mathcal{P}^{i+}_L \), then the price path will stay in \( \mathcal{P}^{i+}_H \cap \mathcal{P}^{i+}_L \) and converge.
- **Case 2.** If \( p^t \in \mathcal{P}^2 \setminus (\mathcal{P}^{i+}_H \cup \mathcal{P}^{i+}_L) \), the price path will either remain within \( \mathcal{P}^2 \setminus (\mathcal{P}^{i+}_H \cup \mathcal{P}^{i+}_L) \) and converge therein, or transition into \( \mathcal{P}^{i+}_H \cup \mathcal{P}^{i+}_L \) (Case 1 and Case 3).
- **Case 3.** If \( p^t \in (\mathcal{P}^{i+}_H \cup \mathcal{P}^{i+}_L) \setminus (\mathcal{P}^{i+}_H \cap \mathcal{P}^{i+}_L) \), then the price path will converge monotonically.

These three cases collectively establish the convergence of the C-OPGA. We complete the proof by demonstrating that the limiting point of Algorithm 2 is an \( \epsilon \)-SNE given that \( \epsilon \) and \( \epsilon_1 \) satisfy the relation in Eq. (18).

Finally, we comment that the convergence result in Theorem 4 is asymptotic, and it is straightforward to derive that the associated dynamic regret satisfies \( \lim_{T \to \infty} \text{D-Regret}_i(T)/T = O(\epsilon) \). This is weaker than the non-asymptotic convergence established in Theorem 2 for the loss-neutral scenario. We speculate the reasons to be two-folded. First, asymmetric reference effects result in non-smooth revenue functions, and it is typical for non-smooth problems to exhibit a slower convergence (Bertsekas 1997). Second, in the loss-averse case where the SNEs are no longer unique, the sub-gradients in different iterations may point toward different SNEs, leading to potential oscillations in the price and reference price trajectories (an example of severe oscillations is illustrated in Figure 3a).

6. Extensions

In previous sections, we assume that each firm \( i \) possesses accurate information regarding its sensitivity parameters \( (b_i, c_i^+, c_i^-) \) as well as its realized market share \( d_i^t \) in each period. This essentially implies that the firm has access to an exact first-order oracle. Yet, a more practical setting worth considering is that the firm can only obtain a rough approximation of the overall market size and needs to estimate the sensitivities from historical data. This would bring extra error to the computation of the first-order derivative, i.e., \( D_i^t \) in Eq. (8) for the loss-neutral scenario, or \( D_i^{t+} \) and \( D_i^{t-} \).
in Eq. (16) for the loss-averse scenario. In this section, we first assess the feasibility of estimating the market share and sensitivities from real data. Then, we proceed to examine the impact of an inexact first-order oracle on the convergence results of the OPGA and the C-OPGA algorithms.

We consider the approximation of market share and the calibration of sensitivities under both uncensored and censored data. In particular, the uncensored data refers to situations where we have access to both purchase and no-purchase information. A salient real-world application arises when two firms offer substitutable products on third-party online retailers, such as Amazon. These platforms can track non-purchase metrics by monitoring users who browse the site without finalizing a transaction. In such an uncensored scenario where the overall market size can be directly observed, the firms can easily determine their market shares, as the sales quantities are accessible through the feedback mechanism. Additionally, the sensitivity parameters can be calibrated via the classical maximum likelihood estimation (MLE) method, with prominent algorithms including Newton-Raphson, BHHH-2 (Berndt et al. 1974), and the steepest ascent. Given the concavity of the log-likelihood function in the MNL model, these numerical algorithms guarantee improvement at each iteration and lead to fast convergence (Wang and Wang 2017, Train 2009).

While the uncensored data provides the full spectrum of customer interactions, firms in many real-world settings must deal with the censored data, where the non-purchases are not readily observable. In the case of censored data, the sensitivities and the total market size in the MNL model can be approximated using the generalized expectation-maximization (GEM) gradient method proposed by Wang and Wang (2017). This is an iterative algorithm that draws inspiration from the expectation-maximization (EM) approach (Dempster et al. 1977, McLachlan and Krishnan 2007). The capability of estimating the sensitivity parameters, the market share, and thereby the gradient $D^t_i$ (or $D^{t,+}_i, D^{t,-}_i$) strengthens the robustness and practicability of the OPGA (or C-OPGA) algorithm, suggesting its great potential for extensive applications in retailing.

In the presence of approximation errors, firms cannot precisely compute the first-order derivative, which renders the convergence results in Sections 4 and 5 not directly applicable. Indeed, if the errors are disruptive enough, they could impede the convergence of Algorithms 1 and 2. However, if the errors are uniformly bounded by some small threshold $\delta$, the following theorem demonstrates the convergence of both price and reference price paths to a $O(\delta)$-neighborhood of the SNE(s).

**Theorem 5 (Convergence Under Inexact First-order Oracle).** Suppose that the firms can only access an inexact first-order oracle such that the errors are uniformly bounded by any $\delta > 0$. Let the step-sizes $\{\eta^t\}_{t \geq 0}$ be a non-increasing sequence such that $\lim_{t \to \infty} \eta^t = 0$ and $\sum_{t = 0}^{\infty} \eta^t = \infty$ hold. Then, the price paths and reference price paths generated by Algorithm 1 (or Algorithm 2) converge to an $O(\delta)$-neighborhood of the unique SNE in the loss-neutral scenario (or the set of SNEs in the loss-averse scenario).
We remark that the inexact first-order oracle studied in our work is different from the stochastic gradient, which generally assumes a zero-mean noise with finite variance. In stochastic gradient case, it is possible to derive the convergence to a limiting point in expectation or with high probability. By contrast, in our case, the noise in the first-order derivative is a kind of approximation error without any distributional properties. Thus, under the step-sizes specified in Theorem 5, we expect the price and reference price paths to approach the neighborhood of the SNE(s) but continue to fluctuate around that area without admitting a limiting point.

The proof of Theorem 5 builds upon those of Theorems 1 and 4. Specifically, we demonstrate that the inexact gradient effectively guides the price path toward the SNE(s) if the magnitude of the true gradient dominates the error. Conversely, if error levels are comparable with the true gradient, we show that the price path should already be close to the SNE(s). Since the step-sizes decrease to zero, Lemma EC.2 implies that the reference price also converges to the $O(\delta)$-neighborhood. The complete proof of Theorem 5 can be found in Appendix H.

7. Conclusion
Despite the growing attention to reference effects within the operations management community, this well-established consumer behavior remains relatively unexplored in competitive frameworks, particularly within partial information settings. Our paper bridges this gap by examining the realm of price competition in an opaque market, taking into account potentially asymmetric reference effects. The problem is structured as an online game with an underlying dynamic state, known as the reference price. We design no-regret algorithms—specifically, the OPGA and its variant, the C-OPGA—and provide theoretical guarantees for their global last-iterate convergence to corresponding SNE(s) in both loss-neutral and loss-averse scenarios. This approach allows the firms to simultaneously achieve the market equilibrium and stability in the long run, even when they do not possess any information about their competitors. Moreover, with loss-neutral reference effects, we show that the OPGA algorithm has the convergence rate of $O(1/t)$ and attains a dynamic regret of $O(\sqrt{T})$ over $T$ periods, given appropriate step-sizes. We further demonstrate that our convergence results can be extended to scenarios with inexact gradients, under which the price and reference price approach the neighborhood of the SNE(s).

Our model gives implications to real-world competitions, particularly within the fast-moving consumer goods (FMCG) industry. Specifically, the opaque market setup in our model parallels the common issue of non-transparency in competitive markets, characterized by the firms’ reluctance to disclose proprietary information to rivals. In addition, the reference effect is non-negligible for the FMCG, as repeated purchases of the FMCG motivate strategic customers to rely on a product’s historical price when making purchase decisions. We propose no-regret learning algorithms
applicable to this practical model, providing firms with simple and effective pricing strategies for reaching the long-term stability and equilibrium without compromising their privacy. Such a stable market enhances predictability, a crucial element for effective supply chain management and long-term business planning, as it fosters firms to make informed decisions and optimize their resource allocation.

This work paves the way for several exciting future research directions. First, this paper unfolds within a deterministic strategy profile, i.e., considering only the pure strategy Nash equilibrium, which is a more common notion found in the literature and proves to be more practical and straightforward for firms to implement. However, it is also worthwhile for future exploration to investigate mixed-strategy learning (Perkins et al. 2015) in dynamic competition problems. Beyond the scope of this work, one interesting direction is to consider other reference price models, for example, the stimulus-based (or external) reference price. This perspective suggests that the price judgment is formed at the point of purchase by using the current external information such as the price of other products (Lynch Jr and Srull 1982, Biehal and Chakravarti 1983, Hardie et al. 1993). Finally, as consumers engage in repeated interactions with retailers, they may gain insights into the pricing strategies and behave strategically in response. Incorporating this forward-looking behavior into the pricing policies presents an intriguing yet challenging opportunity to design more robust pricing strategies.

References


Guo MA, Shen ZJM (2023) Duopoly price competition with reference effects under logit demand. Available at SSRN 4389003.


Ji S, Yang Y, Shi C (2023) Online learning and pricing for multiple products with reference price effects. Available at SSRN 4349904.


