Promotional Strategies on Online Platforms

Zibo Liu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Yong Tan, Chair

Shi Chen

Stephanie Lee

Program Authorized to Offer Degree:

Business Administration

University of Washington

Abstract

Promotional Strategies on Online Platforms

Zibo Liu

Chair of the Supervisory Committee:
Michael G. Foster Endowed Professor, Yong Tan
Information Systems and Operations Management

This dissertation studies various promotional strategies on online platforms by investigating their empirical impact and comparing different policies by analytic models. In three different context, product sampling on e-commerce platforms, online sales with limited inventory, and sampling campaigns on crowdfunding platforms, I delve into the mechanisms behind these promotional strategies to derive reliable explanations on their effects. In my first essay, I study the signaling effect of sampling size in physical goods sampling via online channels. By building a novel structural model, I find that sampling size serves as a positive signal of product quality, and the signaling effect moderated by product types. Further simulation on sampling threshold can help the platform control the scale of the sampling campaign. In my second essay, I study inventory disclosure policies for online sales with limited inventory. I compare the performance of several

existing policies and propose a new policy which is optimal under mild conditions. I also simulate relative improvements of these policies in the numerical study. In the third essay, I study the sampling campaign in crowdfunding markets. Leveraging a structural model, I find that sampling campaign raises the awareness of more backers on the focal project but lowers backers' utility of supporting the project in later stages of the campaign. I also analyze the effect of several signals in the campaign and discuss the aggregated impact of sampling on demand in each stage of the sampling campaign.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Professor Yong Tan. He has guided me all the way through my Ph.D. study and provided me with consistent support. His enthusiasm for research is admirable and always motivate me when I need incentives. His wisdom in methodologies and acute insight in research topics impressed me significantly, and I have learnt a lot from him. He cares about all his students and helped me overcome various difficulties I face as a doctoral student. He helped me schedule my academic career and gave me enough freedom to make my own decisions. I have always felt fortunate to have him as my academic advisor.

My thanks also go to the other committee members, Professor Shi Chen, Professor Stephanie Lee, Professor Ming Fan, and Professor Yuya Takahashi. Their insightful comments and constructive suggestions have helped me to improve this dissertation.

I would also like to express my appreciation to Shawna M. Reimers, Beau Kirkeby, Nuzulita Budhiari, Jaime Banaag, Jessica Aceves, and other staff at the Foster School of Business for their consistent support during my study. My gratitude extends to all fellow doctoral students who have shared this journey.

I thank my wife, Zhen Fang, for her unconditional love and consistent company. No matter what difficulties I encounter, she has always been by my side. I would also like to thank my parents, Jun Liu and Dongxia Qu, for their support through my study. They have always been there to provide me whatever they can provide. It would be hard for me to go this far without the support from my family.

Chapter 1. INTRODUCTION

Online market is growing fast in recent years with the development of various online platforms. To attract more people, online platforms design and implement different promotional strategies. However, without careful examinations on these policies, it is difficult to find out their true impacts or compare the performance of different policies. In my studies, I analyze three different promotional strategies on online platforms, which are product sampling on e-commerce platforms, inventory disclosure policies on online sales with limited inventory, and sampling campaigns on crowdfunding platforms.

In my essays, I use solid methods to improve the reliability of my results. Specifically, I build models to characterize the mechanisms on online markets, and it also helps me to make explanations on my findings more convincible. In my first and third essays, I develop structural models to perform causal analysis in empirical works. In my second essay, I build the analytical model based on queueing theories. Leveraging these methodologies, my studies focus on deriving the true influence of various strategies, providing reliable explanations, comparing existing policies, and proposing new strategies.

In Chapter 3, I study the signaling effect of sampling size in the context of physical goods sampling on online platforms. Prior literature has focused primarily on offline sampling of physical goods and online sampling of information goods, whereas a new form of sampling campaign—physical goods sampling via online channels—receives little attention. Leveraging a rich dataset from Taobao, a large e-commerce platform in China, this essay strives to understand the mechanisms of physical goods sampling via online channels. Specifically, I investigate (1) how the sampling size (number of free samples provided) acts as a signal of product quality and (2) how the signaling effect of sampling size varies across product types (i.e., search, experience, and credence products). I build a structural model to characterize the e-tailers' decision process as a trade-off between the incurred cost of free sampling at present and the economic returns in the future. I find that a 1% increase in sampling size leads to a total sales increase of 5.34% daily sales over a month. Of all product types, experience products benefit the most from the signaling effect of sampling size, whereas search products benefit the least. I further perform a policy simulation on the sampling threshold, that is, the minimum total value of the sampling product required to join the sampling campaign. The results show that increasing the threshold leads to decreases in average sales and the number of e-tailers involved, and vice versa. As one of the first studies focusing on the mechanisms of physical goods sampling in the online context, this essay provides empirical evidence of the signaling effect of sampling size.

In Chapter 4, I study the inventory disclosure policies in online sales of limited inventory. Online sales of limited inventory such as lightning and flash deals have become popular among e-commerce platforms, including Amazon and eBay. Motivated by empirical studies of the impact of inventory information on demand in flash sales (Cui et al. 2019 and Calvo et al. 2020), I study the platform's best timing of disclosing inventory information to maximize the expected sales in a finite horizon. I analyze the following common policies in practice: "always disclose," "never disclose," and the fixed threshold policy. The fixed threshold policy requires the platform to broadcast once the inventory level drops below a preset level. I also propose a time-dependent threshold policy, which requires a comparison between the current time and a time-threshold associated with the current inventory level, and show that the proposed policy is the optimal policy under certain assumptions that are consistent with the empirical findings of the extant literature. For both threshold policies, I devise efficient algorithms to optimize the policy parameters, and I compare them through a numerical study. I find that both threshold policies can significantly outperform the two simple policies of always or never disclose. In particular, when the observational learning (herding) effect dominates the scarcity effect on demand, the fixed threshold policy is near optimal. However, when the scarcity effect dominates, employment of a fixed threshold policy may backfire as the customers may interpret not disclosing as a signal of slow sales. Furthermore, when both effects are not profound, the proposed time-dependent threshold policy can significantly improve the fixed threshold policy, so the platform should employ the proposed policy instead of the fixed-threshold policy. Therefore, this study provides not only effective and efficient algorithms for policy optimization but also guidelines for policy selection.

In Chapter 5, I study the impact of sampling campaigns on crowdfunding markets. The crowdfunding market has developed fast in recent years. However, the problem of information asymmetry in the market is still an issue for both backers and fundraisers. Hence, crowdfunding platforms design various mechanisms to reveal more project information to backers. This essay studies the impact of a novel mechanism in the crowdfunding market, product sampling, on crowdfunding projects. Leveraging a rich data set from a large online crowdfunding platform in China, I construct a structural model considering both the demand and supply sides of the market. I use a consideration set model to capture that the backer cannot notice all projects on the crowdfunding platform. I find that the sampling campaign raises the awareness of more backers on the focal project in all stages but lowers their utility of supporting the project in the latter two stages of the campaign, namely *Trial Stage* and *Report Stage*. Backers’ strategic delay waiting for valuable feedback reports and the negative impact of revealed information on projects can explain the utility drop in these two stages. I also find that the feedback report’s score and the number of free samples offered serve as effective quality signals of projects. Moreover, the counterfactual analysis shows that the demand of most projects soars in *Application Stage* but falls in *Trial Stage*, and half of the projects see a demand increase in *Report Stage* while the other half experience a demand decrease. This research fills the research gap of product sampling in the crowdfunding market, makes methodological contributions, and provides meaningful managerial implications to fundraisers and crowdfunding platforms.

Chapter 2. LITERATURE REVIEW

This dissertation is related with several streams of literature: product sampling, signaling effect, product type, customer's strategic behavior, revenue management, inventory disclosure and inventory management, and information asymmetry in online crowdfunding. I introduce the related literature in detail in following subsections.

2.1 PRODUCT SAMPLING

My first and third essays are both related with the literature of product sampling. Product sampling is a very effective promotional marketing strategy in boosting consumer demand with a long history. Sampling campaigns can be held in different marketplaces, such as offline market (Jain et al. 1995; Schultz et al. 1998; Heiman et al. 2001; Bawa and Shoemaker 2004), online market of information goods (Chellappa and Shivendu 2005; Cheng and Liu 2012; Lee and Tan 2013), and online market of physical goods (Lin et al. 2019). Customers can directly try the free samples (e.g., food and drinks) in an offline market, use a free version of information goods (e.g., software, movie, and music) with limited functionality or limited time in an online market of information goods, and apply for the chance to receive free samples in an online market of physical goods.

Traditional product sampling of physical goods is a very effective yet costly promotional strategy (Jain et al. 1995). It helps consumers learn about a product through direct experience to mitigate perceived risks in product quality (Jamieson and Bass 1989; Rothschild and Gaidis 1981). Previous studies on the traditional offline sampling of physical goods focus on developing optimal sampling strategies (Heiman et al. 2001; Jain et al. 1995) and finding empirical evidence of the effect of product sampling (Bawa and Shoemaker 2004; Kempf and Smith 1998).

Recently, many studies investigate emerging promotional strategies that resemble offline physical goods sampling, such as showrooming (Balakrishnan et al. 2014; Bell et al. 2015; Gao and Su 2017; Letizia et al. 2018) and prefunding (Wei et al. 2021). Showrooming, the behavior of trying the product in brick-and-mortar stores but buying online, is first introduced as a customer strategy (Balakrishnan et al. 2014). Omnichannel retailers soon take advantage of showrooming to lower customers' product fit uncertainty and boost their sales (Bell et al. 2015). Besides physical showrooming, retailers can also leverage virtual showrooming to increase conversion rates and order value (Gallino and Moreno 2018; Letizia et al. 2018). Prefunding, an innovative feature in crowdfunding, enables fundraisers to share project information with potential backers before the fundraising starts. The information that backers receive at the prefunding stage helps them assess the quality of the project more accurately, and projects with prefunding are more likely to succeed (Wei et al. 2021). In offline physical goods sampling, showrooming, and prefunding, customers directly experience the good or acquire the information directly from the product supplier before buying it, whereas customers can see only the amount of samples provided and feedback reports as product signals in the context of my first essay. This differentiates it from this stream of literature.

Another research stream related to the first essay is the study of online information goods sampling. Sellers often provide free trials for information goods that are either time-limited or functionality-limited (Cheng and Liu 2012; Cheng and Tang 2010). Many studies at the OM-IS interface have built analytical models to explore the optimal strategies to offer free trials (Chellappa and Shivendu 2005; Cheng and Liu 2012; Cheng et al. 2015; Cheng and Tang 2010). For example, Mehra and Saha (2018) investigate firms' optimal strategies for launching a new software product by introducing public betas prior to the launch and/or offering free trials along with the product. Cheng and Tang (2010) examine when the firm should introduce the free version, and Cheng et al. (2015) compare three different software free trial strategies. In addition, a growing number of empirical studies quantify the effect of free trials in the consumer software market (Lee and Tan 2013) and mobile app market (Liu et al. 2014). Wang and Özkan-Seely (2018) explore the signaling effect of trial length for information goods sampling and find that high-quality firms offer longer trial periods. In the context of my first study, the mechanism that larger sampling size signals higher quality is analogous to their study. Still, as the method of physical goods sampling is very different from that of information goods sampling, my study diverges from this stream of literature.

My first essay studies physical goods sampling via online channels, which was born to alleviate customers’ concerns about product quality on online platforms. Many studies document various ways for consumers to acquire more information on products, such as online reviews (Chen and Xie 2008; Dellarocas 2003) and e-commerce advertisements (Tan and Mookerjee 2005). My study is most closely related to that of Lin et al. (2019). Lin et al. (2019) also study the role of free product sampling on online platforms. They empirically analyze how a product’s engagement in product sampling affects the product’s review rating. They find that engaging in free product sampling causes an upward bias in product rating. My work extends their findings on the impact of product sampling on online platforms by studying the signaling effect of sampling size on product sales. I also contribute to the literature by studying an innovative sampling method related to the existing ones documented by previous literature. With the prevalence of e-commerce, my research offers critical managerial implications to online retailers, consumers, and platform managers.

Although the product sampling on crowdfunding platforms takes a similar form as the online sampling of physical goods, it can be substantially influential for backers compared to the e-commerce context. On e-commerce platforms, consumers can obtain quality information from other consumers' reviews and word-of-mouth on social media, which are based on the real experience of the products (Chevalier and Mayzlin 2006; Duan et al. 2008; Chen and Xie 2008). The feedback report is yet another piece of consumer review but with rich details. However, backers can only refer to the information provided by the fundraisers and the observational learning of other backers' decisions, which lacks project quality information because no one can obtain the product before a successful funding process. Consequently, two potential outcomes appear. On the one hand, it can intensify the signaling effect of product sampling. Only founders who are highly confident with their products will choose to risk allowing backers to give feedback, encouraging backers to support. On the other hand, backers may consider the feedback reports as a crucial information source and strategically delay their support decisions until the report releases (Levin et al. 2009; Yu et al. 2016; Papanastasiou and Savva 2017), hindering the progress of fundraising and leading to potentially unfavorable outcomes (Bøg et al. 2012; Burtch et al. 2013; Koning and Model 2013; Kuppuswamy and Bayus 2018). As the crowdfunding market is quite different from the markets I mention above, it is necessary to analyze the mechanism and investigate the effect of product sampling campaigns in this market.

2.2 Signaling Effect

My first and third essays are also related to the literature of signaling effect. Signaling has been investigated as a means to address information asymmetry in the literature of economics, marketing, OM, and IS since the initial development of signaling theory by Spence (1973). The design and evaluation of various information structures as innovative quality signals have received continuous attention (e.g., Wang and Özkan-Seely 2018; Wei et al. 2021). In the e-commerce context, consumers cannot directly observe product quality when making purchase decisions due to the time lag of product delivery (Mavlanova et al. 2012). Therefore, the asymmetric information between consumers and sellers causes market inefficiency (Akerlof 1970). Fortunately, the concern about information asymmetry can be mitigated by providing quality-related signals in online commerce (Kirmani and Rao 2000; Wells et al. 2011). There are three types of signals in the e-commerce context, that is, signals provided by previous consumers, platforms, and sellers.

As to signals provided by previous consumers, online reviews are considered one of the most important information sources to assist consumers' decision making and thus have been well studied (Chevalier and Mayzlin 2006; Dellarocas 2003; Duan et al. 2008). The volume and the valence of online reviews serve as typical signals of product quality (Chevalier and Mayzlin 2006; Duan et al. 2008). Abrahams et al. (2015) propose a text analytic framework to find product defects using signal cues extracted from user-generated content such as online reviews. Apart from influencing consumers' purchase behaviors, online reviews also exert an influence on other consumers' reviewing behaviors (Wang et al. 2019). However, as online reviews can be easily influenced by various factors that affect consumer behaviors, they may not be an ideal signal in e-commerce decision making (Dellarocas 2006; Ho et al. 2017; Lee et al. 2018). Besides previous customers, online platforms also strive to provide new designs of information structure as a signal to facilitate consumers' decision making. However, platforms can suffer from a certain extent of information asymmetry, and the signals they provide may contain biased information (Ghose et al. 2014; Meng et al. 2020). Finally, sellers are also motivated to provide additional information and even pay for it because they want to distinguish themselves from others leveraging high-quality signals (Kihlstrom and Riordan 1984; Milgrom and Roberts 1986; Nelson 1974). According to prior OM and IS studies, there are various well-understood signals from the supply-side on e-commerce platforms, including discount (Cao et al. 2018), live chat (Tan et al. 2019), and review response (Kumar et al. 2018; Yang et al. 2019). Researchers have also compared the differential effects from signals provided by different parties such as sellers' coupon information versus consumers' online reviews (Ba et al. 2020) and providers' service outcomes versus service clients' reviews (Sun and Xu 2018).

Because reviews act as important signals to consumers, some sellers offer rewards to consumer reviews when they expect to receive positive feedback thanks to high product quality (Li et al. 2020). Those reviews induced by sellers are called incentivized reviews, which are often biased due to firms' manipulation (Luca and Zervas 2016; Mayzlin et al. 2014) or platforms' monetary incentives to stimulate reviews (Khern-am-nuai et al. 2018). In my first essay, the sampling size on online platforms undertakes similar mechanisms as above. E-tailers will participate in product sampling only when two conditions are met: first, the seller can gain a high profit to afford the cost of product sampling, and second, the expected scores of the feedback reports are high. However, the mechanisms between incentive reviews and my study are different. By examining the impact of sampling size, I argue that sampling size serves as an effective signal of product and as another critical information source for consumers.

Regarding the quality of crowdfunding platforms, prior studies suggest that project target (Chakraborty and Swinney 2021), pre-order price (Sayedi and Baghaie 2017), and third-party endorsement (Courtney et al. 2017) can become effective signals. Related literature also reveals that reward supply, social ties, the presentation format of project information, and communication between backers and the founder are also important signals in the crowdfunding process (Kunz et al. 2017). In the context of my third essay, feedback reports play a similar role as online reviews on e-commerce platforms, which is the most critical signal and has been well studied by previous researchers (Chevalier and Mayzlin 2006; Duan et al. 2008). The ratings of online reviews and the volume of online reviews both serve as typical signals of product quality. Thus, I also expect the feedback report score to be a crucial signal of project quality. In terms of the number of samples provided, Liu et al. (2021) identify its signaling effect in sampling campaigns on e-commerce platforms. Following the same logic, the number of samples is also expected to be an effective signal of project quality in crowdfunding markets.

2.3 PRODUCT TYPE

The literature of product type is related with my first essay. The previous economics and marketing literature have classified products into three types (i.e., search, experience, credence) (Darby and Karni 1973; Nelson 1970). Based on prior definitions, search products have standardized attributes (e.g., the memory size of a computer, screen resolution of a mobile phone) for consumers to infer their quality before usage/sampling. The quality of experience products is difficult to assess before usage/sampling, but it can be understood after usage/sampling, whereas the quality of credence products is difficult to evaluate both before and after usage/sampling.

In the traditional offline context, only experience goods appear in sampling campaigns (Klein 1998; Wan et al. 2012). Wan et al. (2012) argue that consumers use direct inspections to evaluate search goods, and they rely on brand names and recommendations to purchase credence goods. They assert that sampling is suitable only for experience goods. However, all types of products can participate in sampling campaigns on online platforms and potentially benefit from them. Prior studies have shown that different types of products face different impacts of information disclosure (Darby and Karni 1973). Thus, in the context of physical goods sampling via online channels, consumers may rely the least on sampling campaigns to infer product quality for search goods, whereas the contrary is true for credence goods. In other words, we expect that among the three product types, the impact of sampling size on product sales would be the weakest for search products and the strongest for credence products. My study enriches the theoretical dialogue on the signaling effect on e-commerce platforms contingent on product types. I demonstrate that the signaling effect of sampling size varies across product types and provides practical guidance to e-tailers on picking out appropriate products for sampling.

2.4 Customer's Strategic Behavior

My second and third essays are related with this stream of literature. In the case of my second essay, in anticipation of customers' strategic reactions, a firm can leverage its operational strategies, including pricing and inventory replenishment decisions, to maximize its revenue or net profit, which leads to a body of literature on operations management in the presence of strategic customers. Previous studies in this literature suggested that the firm can employ a quick response strategy to mitigate customers' strategic behavior (Cachon and Swinney 2011 and Aviv et al. 2019), whereas the firm's inventory pooling strategy may encourage more customers to delay their purchasing decisions, thus aggravating their strategic behavior (Aflaki and Swinney 2019). Yu et al. (2016) examined a firm's pricing strategy and consumers' purchasing behavior in the presence of consumer- generated quality information, which they found is not always beneficial to the firm and customers. Papanastasiou and Savva (2017) also examined a firm's pricing strategy in the presence of social learning (e.g., consumer reviews), and they found that a pre-announced pricing plan is not as good as responsive pricing, which is different from the conventional wisdom without social learning.

My third essay is related to the stream of literature on customers' strategic delay in decision-making. Previous studies show that customers may delay their decision-making on purchasing products when they need more product information (from both the producer and other customers) or when they expect the price to fall (Greenleaf and Lehmann 1995; Cachon and Swinney 2011). Many studies have been conducted in the area of operations management on what dynamic pricing strategy firms should employ when customers can strategically delay their purchasing behavior (Levin et al. 2009; Yu et al. 2016; Papanastasiou and Savva 2017). Similarly, product sampling could negatively impact projects if backers are longing for the feedback reports and refuse to make the decision before the reports appear. Consequently, the low performance of the early funding stage can hurt the performance of the subsequent funding stage, and backers may hold back or withdraw their contributions (Bøg et al. 2012; Burtch et al. 2013; Koning and Model 2013; Kuppuswamy and Bayus 2018). Therefore, the effect of the sampling campaign becomes more complex and requires careful empirical analyses.

2.5 Revenue Management

My second essay contributes to an emerging body of literature on revenue management for online retail platforms. For a general review of revenue management, see Talluri and Van Ryzin (2006) or Özer and Phillips (2012). In particular, my study focuses on the context of online sales of limited inventory such as lightning and flash deals. Ferreira et al. (2016) examined a dynamic pricing strategy for a platform using machine learning techniques. Ferreira et al. (2018) devised dynamic pricing plans to address a tradeoff between learning demand by offering different prices and maximizing the expected revenue. Sodero and Rabinovich (2017) developed a demand forecasting model based on the Bass diffusion model to predict the stockout times of flash sales.

I study the platform's revenue management problem with limited inventory from a different angle: the decision on when to disclose the inventory. My work is most related to Cui et al. (2019) and Calvo et al. (2020). The former reported evidence from Amazon lightning deals that revealing low inventory levels can stimulate future demand, and the latter confirmed the positive effects of the disclosure of limited inventory on sales. Note that the inventory disclosure policies examined by these two studies are different. In Amazon lightning deals, the platform broadcast its inventory from the beginning, whereas the platform studied by Calvo et al. (2020) employed a fixed threshold policy. Although the above two studies have confirmed the causal effects of disclosing inventory on demand, they did not analyze conditions under which a policy is better than the other. My study complements this body of literature by optimizing a platform's inventory disclosure decision.

2.6 INVENTORY DISCLOSURE AND INVENTORY MANAGEMENT

This part of literature is related to my second essay. Although my study is somewhat related to the literature on operations management faced with strategic customers, I focus on the firm's policies of information disclosure, particularly the inventory availability information. Yin et al. (2009) examined how to mitigate customers' strategic behavior through changing the inventory display format. They considered two display formats, display all (DA) and display one at a time (DO), and confirmed that the latter increases the customers' propensity to buy, but they did not consider optimizing the inventory disclosure time as in my paper. Küçükgül et al. (2021) investigated the optimal information policy for revenue-maximizing platforms with unlimited inventory under a dynamic Bayesian persuasion framework. In their context, customers can update their belief based on their private signal and the message that the platform shows to them; they proposed that the platform should provide neutral recommendations to early customers and only affirmative recommendations to customers after a time threshold. In contrast, my work focuses on the platform's strategy regarding when to disclose the inventory availability information (rather than recommendations) for online sales of limited inventory (rather than unlimited inventory), which is different from their setting.

Last, from the modeling perspective, my second essay is related to a body of literature on inventory management where a firm sells a seasonal product and the timing of its decision (e.g., when to place an order) plays an important role. For example, Li and Kouvelis (1999) considered a firm's procurement strategy in an environment where the input price of raw materials fluctuates while the demand is deterministic. The firm monitors the change of the input price and determines the optimal time of purchasing the materials. Other similar studies include Wang and Tomlin (2009), Martínez-de-Albéniz (2011), and Chen et al. (2021). My work is related to these studies in the sense that the firm faces a "wait-or-not" decision as the time progresses. However, there are obvious differences between the above studies and mine, as this study focuses on the firm's inventory information provision policies rather than its inventory replenishment strategies. Moreover, in my study, the firm's decision on when to disclose its inventory can significantly affect customers' reactions and the future demand, whereas, the firm in the above studies cannot.

2.7 Information Asymmetry in Online Crowdfunding

This stream of literature is related to my third essay. Crowdfunding platforms can be categorized into four types: reward-based, equity-based, donation-based, and lending-based (Hemer 2011). My research focuses on reward-based crowdfunding, where founders are allowed to conduct product development and demand test before committing extensive investments (Belleflamme et al. 2014; Wei et al. 2021). In exchange, backers can receive different levels of rewards, usually different versions of products produced in the future, according to their contribution levels.

The information shared with backers is usually limited due to the lack of information disclosure regulations on reward-based platforms and founders' protection of intellectual property from their competitors (Agrawal et al. 2014). Therefore, the high uncertainty in project quality discourages backers from contributing and causes market inefficiencies and failures on the platforms or so-called the information asymmetry problem (Akerlof 1970). On the one hand, to overcome the difficulty, backers strive to seek more information on crowdfunding projects, relying on the choice of others, which is known as the herding effect. Many researchers analyze the herding effect in crowdfunding and other similar online markets and the factors that may affect it (Zhang and Liu 2012; Burtch et al. 2013; Jiang et al. 2018; Burtch et al. 2018). Zhang and Liu (2012) distinguish the rational herding (observational learning), inferring the creditworthiness of founders by observing others' decisions, with the irrational one, passively mimicking others. Jiang et al. (2018) empirically show that the herding effect can be strengthened by the platforms' market share and the cumulative funding amount but be attenuated by the operation time of the platforms. Burtch et al. (2018) show that, compared to keep-it-all, the all-or-nothing scheme of crowdfunding platforms can weaken the herding effect, because it eliminates the concerns of partial fundraising, and the fundraising target signals a high quality of the project.

On the other hand, crowdfunding platforms also pay much attention to the information asymmetry problem. They try to alleviate backers' concerns by designing various policies. For example, Zhou et al. (2021) investigate the impact of loan guarantors, usually large-scale enterprises, and find that their existence triggers the learning process of investors and can moderate the investors' responses to listing attributes in crowdfunded supply chain finance markets. Wei et al. (2021) examine prefunding, an information control policy on reward-based crowdfunding platforms, which enables founders to share information with potential backers before the project starts. The results show that prefunding can democratize the funding outcomes and benefit projects with a less entrepreneurial endowment. Among these policies, product sampling is a novel one in the crowdfunding market and has received little attention in academia. This policy can persuade backers to contribute using the extra information offered by the feedback reports from successful applicants of the free samples. Besides, participating in product sampling signals a high quality of the project and the confidence of fundraisers. However, prior studies also suggest more information disclosure can reduce the magnitude of the herding effect, leading to potentially unfavorable outcomes for fundraisers (Jiang et al. 2018; Zhang and Liu 2012). My study analyzes the impact of sampling campaigns on crowdfunding and provides valuable advice for fundraisers and crowdfunding platforms.

Chapter 3. THE SIGNALING EFFECT OF SAMPLING SIZE IN PHYSICAL GOODS SAMPLING VIA ONLINE CHANNELS

Free product sampling has always been a popular promotional strategy employed by firms to boost product sales (Schultz et al. 1998). In the traditional offline context, marketers in various venues (e.g., retail stores, supermarkets, and shopping malls) often provide free samples of physical goods (e.g., food and drinks) for consumers so that consumers can reduce product quality uncertainty before committing to a purchase (Heiman et al. 2001). Facilitated by the development of the Internet in the past decade, product sampling in brick-and-mortar stores has extended to the online context (Reid 2013). Unlike offline product sampling, though, most sampling products in the online context are information goods (e.g., software, movies, and music).

Recently, however, many electronic commerce (e-commerce) websites (e.g., Taobao.com, JD.com) have introduced their own sampling promotions of physical goods. Some online platforms are now dedicated to helping retailers perform sampling to targeted consumers. These platforms grew rapidly during the Covid pandemic, such as Brandshare connected with over 800 e-commerce retailers (Taylor 2020) and Sampler with over 400 brands (Sampler 2020). The new form of physical goods sampling via online channels is becoming increasingly popular as e-commerce websites burgeon. Specifically, e-commerce retailers offer a certain number of free samples (hereafter termed as sampling size) (Jain et al. 1995) with free delivery for online consumers to apply for. The successful applicants can receive the free samples and keep them after the sampling promotion. In return, they need to write feedback reports on sampling products after the tryout. These feedback reports, usually full-length product evaluations, are displayed separately from typical online reviews.

Albeit widely adopted in industry, the emerging physical goods sampling via online channels is underexplored in academia. Thanks to the nature of online platforms, this new form of product sampling is embedded with different mechanisms from traditional ones. Offline physical goods sampling and online information goods sampling offer either free trials of a small size (e.g., food, drink, etc.) or a limited version of the product (e.g., time-limited or functionality-limited versions of the software) to mitigate the product quality uncertainty via direct experience (Heiman et al. 2001; Cheng and Liu 2012). However, sampling campaigns on online platforms provide complete products and rely on the signaling effect and feedback reports to convince customers of high product quality. The signaling effect stems from the fact that sampling is costly, and only e-tailers with enough confidence in their product quality would take part in the sampling campaign (Spence 1973). Feedback reports can also reduce product uncertainty and help consumers make better decisions (Chevalier and Mayzlin 2006; Duan et al. 2008). As past sampling campaigns are available to all consumers, the sampling effect of physical goods on online platforms is a long-term effect, another distinctive aspect. In addition, physical goods sampling on online platforms allows all types of products to join, whereas sampling campaigns in other contexts offer only experience goods and information goods. The signaling effect is a novel and interesting mechanism in the context of physical goods sampling via online channels, which does not take place in other sampling campaigns. It is essential to understand the unique mechanism behind it, for both academia and industry (the e-tailers, the platform, etc.). Table 3.1 summarizes the differences between our new form of sampling campaign and previous forms.

Table 3.1. Physical Goods Sampling via Online Channels vs. Other Sampling Campaigns

	Physical goods sampling via online channels	**Offline physical goods sampling**	**Online information goods sampling**
Example	A Lenovo laptop sampled on Taobao Try	A new flavor of Lays potato chips sampled in a supermarket	A 30-day free trial of EndNote, the basic version of Grammarly

Product Completeness	A complete product	Free trials of small size (e.g., food, drinks, etc.)	Time-limited or functionality-limited free trials
Product Type	All types	Experience goods	Information goods
Recipient	A limited number of successful applicants	All customers; per request	All customers; per registration
Additional Information	Feedback reports	Direct experience	Direct experience
Mechanisms	Signaling effect & Lower product uncertainty & WOM	Lower product uncertainty & WOM	Lower product uncertainty & WOM
Effect Duration	Long-term	Short-term	Short-term (time-limited) / long-term (functionality-limited)

Notes. Grammarly is an automated grammar checker software. WOM = Word-of-Mouth.

Our paper investigates the signaling effect in physical goods sampling via online channels. Specifically, how does sampling size affect consumer evaluations of product quality as a signal, thus impacting sales? And how does the effect of this signal change across product types? The sampling of physical goods is always considered an expensive promotional strategy (Jain et al. 1995). This claim is also valid in the context of sampling on online platforms. On the one hand, online sampling platforms usually impose eligibility criteria (i.e., sampling threshold, a requirement on the total value of the product) for e-tailers to participate in online sampling campaigns. E-tailers hence must consider the trade-off between the incurred cost of free sampling at present and the economic returns gained in the future. On the other hand, as the cost of sampling physical goods increases with the sampling size, e-tailers face a cost-efficiency challenge: a larger sampling size generates more consumer awareness, but it also bears a higher associated cost. Their decision process is more complicated, as e-tailers also need to consider the effect of the feedback reports' scores on future sales. Lastly, the above considerations are unlikely to be equally applied to all types of products (e.g., search, experience, and credence products). For different types of products, consumers perceive their quality differently. Consequently, product types may affect the effectiveness of sampling size serving as the quality signal. Thus, studying the roles of sampling size and product type in product sampling on online platforms can fill the research gap in the product sampling literature and help us understand the mechanism behind this new form of sampling campaign. It can also provide managerial implications to both e-tailers and platform managers. We thus address these essential issues in this study. In particular, our research questions are:

(1) How does the sampling size of a physical product in online sampling promotion affect the product's sales?

(2) How does the impact on sales depend on product type (i.e., search, experience, and credence products)?

To answer the questions above, we leverage a unique data set from Taobao Try, the online physical goods sampling platform developed by Taobao (the largest e-commerce platform in China). Taobao provides a diverse environment where products sold by e-tailers vary greatly and include all three product types mentioned above. E-tailers can participate in sampling campaigns on Taobao Try by providing several free samples for customers. Taobao Try requires a minimum total value of the sampling product, which means that e-tailers must provide enough free samples to be eligible for a sampling campaign. Given the sampling rules set by the platform, e-tailers make their own decisions on whether to join a sampling campaign and, if so, how many free samples to offer.

In this paper, we build a structural model to capture the decision process of e-tailers regarding the sampling campaign. E-tailers are rational decision makers who make expectations on future sales conditional on various sampling sizes and choose the optimal one. To get an unbiased estimation of the signaling effect, we disentangle it from the impact of feedback reports in our model. To further alleviate the endogeneity issue, we also adopt the idea of BLP (Berry et al. 1995) to characterize e-tailers' decision processes. Our results show that sampling size has a significant and positive signaling effect on e-tailers' daily sales. Specifically, a 1% increase in sampling size leads to a total sales increase of 5.34% daily sales over a month. Of all types of products, experience goods benefit the most from sampling campaigns, whereas search goods benefit the least. Credence goods benefit more than search goods but less than experience goods, which is a result that may go beyond the expectation of e-tailers selling credence goods. E-tailers can learn the effect of sampling size from our results to help them make wise decisions on online sampling campaigns.

We further investigate how e-tailers' decision to join the sampling campaign is affected by the online platform's sampling threshold. Through our policy simulations, we find that the number of sampling shops and average sales both decrease with the sampling threshold, yet their decreasing patterns are slightly different. With the current threshold of ¥1,500, lowering the threshold to ¥200 can raise average sales by 5.83% and encourage 57.2% more shops to participate in the sampling campaign, whereas increasing the threshold to ¥3,000 results in a 6.99% decrease in average sales and blocks 21.2% sampling shops out of the sampling campaign. Platform managers can leverage our simulation results to find a potentially better sampling threshold.

Our study makes considerable contributions to the Operation Management (OM) and Information Systems (IS) literature. First, we fill the research gap on the topic of product sampling in OM literature by studying mechanisms of an emerging scheme in product sampling, that is, physical goods sampling via online channels. While offline physical goods sampling and online information goods sampling rely on directly experiencing sampling products to reduce uncertainty about product quality (Heiman et al. 2001), our study provides empirical evidence on the unique signaling effect of sampling size in physical goods sampling on online platforms. Furthermore, e-tailers can easily leverage our results on the signaling effect of sampling size to better understand product sampling on online platforms.

Second, we extend existing works in the IS and OM literature on signaling effects by identifying and evaluating sampling of physical products via online channels as an innovative quality signaling design on e-commerce platforms. On online platforms, product quality-related signals are introduced to alleviate information asymmetry (Kirmani and Rao 2000; Wells et al. 2011). Previous OM and IS studies have investigated the design and evaluation of various information features as quality signals (e.g., Lee et al. 2018; Wang and Özkan-Seely 2018; Wang et al. 2019) and focus on various signals provided by previous customers (Chevalier and Mayzlin 2006; Dellarocas 2003), sellers (Kihlstrom and Riordan 1984; Milgrom and Roberts 1986) and platforms (Ghose et al. 2014; Meng et al. 2020). Our research adds to this body of literature by identifying the sampling size of the sampling product as a new signal from sellers.

This study also contributes to the literature on the effects of product types. Out of three product types (Darby and Karni 1973), only experience goods can benefit from sampling campaigns in the traditional brick-and-mortar market (Klein 1998; Wan et al. 2012). However, our empirical results indicate that products of all three types can gain from sampling campaigns on online platforms, with experience goods benefitting the most. Our study is the first to document the sampling effect on search goods and credence goods in the online context. Our work also adds to the literature of methodologies used in the online context. We propose a novel structural model to perform an unbiased estimation in physical goods sampling on online platforms. Future work can leverage our model in similar cases where the decision process of customers or sellers is endogenous and not negligible.

3.1 DATA

Our data set was crawled from Taobao.com, which was launched in 2003 by the Alibaba Group, one of the largest e-commerce websites in China (Alexa 2020). Taobao has more than 500 million registered users, 60 million daily active users, and 800 million products (Taobao 2020). Each e-tailer on Taobao has one shop website, and each product of the shop has a separate webpage. On the shop website, consumers can observe shop ratings, shop reviews, and shop products. On the product webpage, consumers can observe product characteristics such as product titles, product reviews, and prices. More importantly, unlike other e-commerce websites, which do not provide product sales information or display only product sales rankings (e.g., Amazon), Taobao released all the detailed transaction records of a product at the time of our data collection. We took advantage of this feature and crawled actual product sales.

Taobao established its sampling platform Taobao Try (i.e., Taobao sampling center, http://try.taobao.com) in 2011. It is now the most prominent online sampling platform in China. The sampling promotion procedure breaks down into several steps. First, the e-tailer who takes part in the sampling campaign decides how many free samples to provide. Taobao Try sets ¥1,500 as the minimum total value of sampling products for e-tailers applying for a sampling campaign. In other words, the e-tailer needs a minimum sampling size to join the sampling campaign. Next, the sampling center approves the request and displays the product (with detailed information such as product title, product price, sampling size, promotion starting date, etc.) in the "Coming Soon" section several days (typically three days) before the start of the actual sampling promotion. Then, when the sampling campaign starts, the product will be moved to the "Ongoing" section, and any registered user of Taobao can apply for a free sample of the sampling product. The sampling campaign lasts for seven days, and the number of applicants is displayed during the sampling period. After the campaign finishes, the sampling center selects successful applicants according to their past behavior on Taobao (number of past sampling product applications, amount of past spending, etc.), and the product is moved to the “Completed” section. The e-tailer then delivers free samples to successful applicants. Finally, successful applicants use the samples and are required to submit feedback reports to the sampling center. These reports will be displayed on the product webpage in the sampling center.

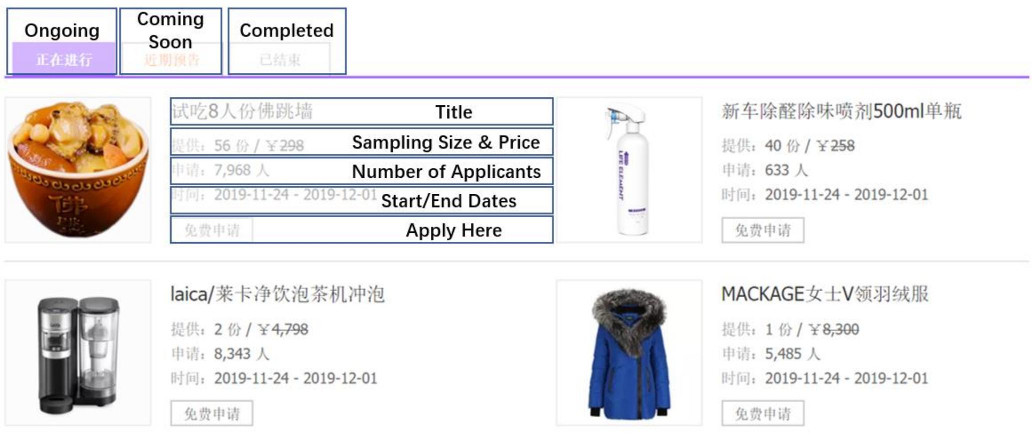

Figure 3.1. Home Page of Taobao Try

Figure 3.2. Webpage of a Sampling Product on Taobao Try

Figures 3.1 and 3.2 show examples of the homepage of Taobao Try and a webpage of a sampling product in the “Ongoing” section on Taobao Try. The Taobao Try home page shows “Ongoing,” “Coming Soon,” and “Completed” sampling campaigns. It also shows detailed information about sampling products, including title, sampling size, regular price, number of applicants, and starting and ending dates of campaigns. Customers can click “Apply Here” and land on the webpage of a specific sampling product. The sampling product webpage provides similar information to that on the home page, except that it contains a product image that links to the product’s webpage on the Taobao platform and an “Apply Here” button for customers who want to apply for free samples after they browse this webpage. Also, the sampling product webpage shows the remaining time of the sampling campaign.

We collect data across a period of seven months on a daily basis. Our data collection follows these steps. First, we check the “Coming Soon” section and add a sampling product and its shop to the crawling list when it first appears in this section. Then, on the same day, we use Taobao’s search function to search with the sampling product title to find other products with almost identical titles. We generate a random number k and use the k-th product from the search results to ensure the randomness of the non-sampling product. We also add this non-sampling product and its shop to our crawling list. The non-sampling product can be quite different from the sampling product, and their product quality may differ widely. Also, as Taobao is such a competitive marketplace, customers attracted by the sampling product are more convinced by its quality than other similar products’. Thus, it is unlikely that there will be a spillover effect of the sampling campaign. Based on the crawling list of products and their shops, the crawler goes through each product webpage each day to gather and record all observable information, such as product and shop attributes, transactions, and sampling promotions.

As the daily crawling task is highly computationally intensive, we separate our data collection into three consecutive phases. We collect a new set of sampling products and non-sampling products during each phase. These sampling products are randomly collected from each category on the platform. To be specific, among all the products in a category, we generate n random numbers and select products with their order corresponding to these numbers. We set the n's of each category to be proportionate to the total number of sampling products of those categories. We collect about 40 sampling products in total (depending on availability) during each phase. Each phase lasts for about a month so that we have a complete observational period of a product, which starts before the start of the product's sampling promotion and ends after receiving the feedback reports. We finally combine these three sets of data to assemble a rich data set. Overall, our final unbalanced panel-level data set includes 7,076 observations for 262 products and their corresponding shops. Based on Taobao's categorization, all the products are grouped into seven categories: (1) computers and accessories, (2) mobile phones and accessories, (3) vacuum cleaners, (4) biscuits, (5) drinks, (6) wine and liquor, and (7) health supplements. Based on the prior literature, we consider products in categories (1) to (3) as search products, those in categories (4) to (6) as experience products, and those in category (7) as credence products.

We show the summary statistics of some variables in our data set in Table 3.2. Product and shop information are all shown on a daily level except sampling size, work hours, and the number of workers. Statistics on sampling size are calculated at the product level on products that participate in sampling campaigns. Statistics on work hours and the number of workers are calculated at the shop level. Size and Size × Price are sampling size and sampling value for sampling products. The sampling size of all products ranges from 1 to 90, with an average of 11.73, and the mean sampling value is more than ¥2,000. Sales stands for the sales quantity of a product. AvgReportScore is the average score of feedback reports for sampling products. We consider ReviewVolume but not ratings for products, as only products in a sub-platform of Taobao (i.e., Tmall platform) have a rating, whereas we can collect the number of reviews for all products.

Delivery indicates whether there is a delivery fee or not. PromiseNum is the number of after-sales promises (e.g., product warranties, free returns) offered to customers on the product website. For shops, we have both ratings as ShopRating and rating volumes as ShopRatingVolume. WorkHour is the number of hours every week that the customer service of the shop is available. WorkerNum is the number of workers contacting customers in the shop. We consider WorkHour and WorkerNum as potential factors correlated with each shop's decision process of choosing the optimal sampling size.

3.2 MODEL

In our context, the effect of sampling size on sales cannot be directly estimated from simple regression, as sampling size is not an exogenous variable. The reason is that shops make decisions on sampling size based on their characteristics, including unobserved variables. As a result, their final choices of sampling size are correlated with their unobserved characteristics. Thus, if we directly estimate the impact of sampling size on sales without any controls of shops' decision process regarding sampling campaigns, we may get a biased estimate.

Table 3.2. Description of Variables

Variable	Mean	Std. Dev.	Min	Median	Max
Product Information					
Sales	10.15	70.37	0	0	2,739
Size	11.73	14.49	1	8	90
Price	306.35	623.32	5.98	99	4,199
Size × Price	2,192.53	1,306.96	1,500	1,680	9,960
AvgReportScore	4.83	0.23	3.25	4.89	5
ReviewVolume	1,539.72	7,468.65	0	52	113,601
Delivery	0.085	0.279	0	0	1
PromiseNum	1.99	1.22	0	2	5
Store Information					
ShopRating	4.82	0.08	4.3	4.8	5
ShopRatingVolume	18,962.74	65,985.01	1	2,470	794,117
WorkHour	117.33	30.40	0	108	168
WorkerNum	3.90	2.26	1	4	17

To solve the endogeneity problem, we develop a structural model. We follow the approach of controlling endogenous price in Berry et al. (1995) and apply a shop's decision model on sampling size. In the decision model, we assume that the shop makes a trade-off between the incurred cost of free sampling at present and the economic returns gained in the future. We also need to disentangle the signaling effect from the impact of the feedback report's score in our model. The latter reflects another critical channel through which a sampling campaign can affect sales. In addition, we use instrumental variables to alleviate the concern of endogenous price.

3.2.1 *Demand Model*

In our demand model, we consider how sales quantity is affected by various factors. On the webpage of a product on Taobao platform, consumers can observe many characteristics of the product and its shop. The webpage shows the name of the product, several pictures of it, price, product reviews, number of product reviews, delivery policy, the shop's selling promises on this product (such as free returns within a week), shop ratings, number of shop ratings, etc. Consumers make purchase decisions based on these pieces of information. As we only have sales data of the product aggregated at a daily level without characteristics of buyers, we use a linear model on sales as follows:

$$\begin{aligned} ln\left(Sales_{jt}\right) = \beta_0 &+ \beta_1\, ln\left(Size_{jt}\right) + \beta_2\, ln\left(Price_{jt}\right) + \beta_3\, ln\left(Price_{jt}\right) \times Sample_j \\ &+\beta_4 HasReport_{jt} + \beta_5 AvgReportScore_{jt} + \beta_6 DaysElapsed_{jt} \\ &+\beta_7 Delivery_{jt} + \beta_8 PromiseNum_{jt} + \beta_9\, ln\left(ReviewVolume_{jt}\right) \\ &+\beta_{10} ShopRating_{jt} + \beta_{11}\, ln\left(ShopRatingVolume_{jt}\right) \\ &+\beta_{12} Type_j + \beta_{13} DayType_t + \xi_j + \varepsilon_{jt} \end{aligned} \tag{3.1}$$

In the above equation, j indicates product, and t indicates day. $Size_{jt}$ is the sampling size of product j at time t. For non-sampling products, $Size_{jt}$ is always 0. For sampling products, $Size_{jt}$ is 0 before the sampling campaign begins and equals the sampling size after the campaign starts (even after the campaign finishes). Since consumers can still observe the information of the sampling campaign on the platform after the campaign finishes, we assume the effect of sampling lasts in the period we observe (about a month) and do not set $Size_{jt}$ back to 0 after the campaign finishes. We add an interaction term between $ln(Price_{jt})$ and the indicator of sampling product $Sample_j$ to partly address the endogeneity issue of price, as the price and report score of a sampling product are correlated. $HasReport_{jt}$ is an indicator of having feedback reports for sampling products. Report variables are set to 0 before the first feedback report of the product releases. $DaysElapsed_{jt}$ is days elapsed since the sampling campaign began, and it captures the diminishing sampling effect. For non-sampling products, this variable is always 0. $Type_j$ are type dummies of product j. $DayType_t$ is the dummy indicating whether day t is weekday, weekend, or holiday. Other variables have the same meanings as those in Table 3.2 and are observed on a product-day level. ξ_j is a product random effect that represents the product's unobserved quality that may affect product sales, and ε_{jt} is an idiosyncratic error. All logarithms of variables are taken after adding one.

To explore heterogeneous effects of sampling size for different product types, we add interaction terms of $ln(Size_{jt})$ and product type dummies as follows:

$$\begin{aligned} ln(Sales_{jt}) = \beta_0 &+ \beta_1\, ln(Size_{jt}) + \beta_2 Search_j \times ln(Size_{jt}) + \beta_3 Credence_j \times ln(Size_{jt}) \\ &+\beta_4\, ln(Price_{jt}) + \beta_5\, ln(Price_{jt}) \times Sample_j + \beta_6 HasReport_{jt} \\ &+\beta_7 AvgReportScore_{jt} + \beta_8 DaysElapsed_{jt} \\ &+\beta_9 Delivery_{jt} + \beta_{10} PromiseNum_{jt} + \beta_{11}\, ln(ReviewVolume_{jt}) \\ &+\beta_{12} ShopRating_{jt} + \beta_{13}\, ln(ShopRatingVolume_{jt}) \\ &+\beta_{14} Type_j + \beta_{15} DayType_t + \xi_j + \varepsilon_{jt} \end{aligned} \tag{3.2}$$

In the sales equation, the endogeneity problem arises from the correlation between sampling size and ξ_j, report score and ξ_j, and price and ξ_j. The correlation of sampling size and ξ_j comes from the fact that the shop decides how many free samples it wants to provide for the sampling campaign. On the one hand, a higher positive random effect ξ_j would induce the shop to expect a larger increase in future sales thanks to the sampling campaign so that the shop would offer more free samples. On the other hand, the random effect ξ_j may be correlated with the shop's trade-off procedure between the loss of offering free samples without receiving any revenue and the gain of future sales from providing free samples. In other words, a shop with unobserved high quality may consider future sales to be of higher or lower importance and thus decide to offer more or fewer free samples. Through these two channels, sampling size becomes correlated with the unobserved quality of the product, ξ_j. Combined with the supply model on the decision process of shops in Section 3.2.2, we can eliminate the endogeneity on sampling size.

As a product of high quality may also achieve high scores from feedback reports, we need to consider the correlation between report score and ξ_j. Thus, we build a model for report score as follows:

$$\begin{aligned} AvgReportScore_j = \gamma_0 &+ \gamma_1 \ln(Price_j) + \gamma_2 Delivery_j + \gamma_3 PromiseNum_j \\ &+ \gamma_4 \ln(ReviewVolume_j) + \gamma_5 ShopRating_j \\ &+ \gamma_6 \ln(ShopRatingVolume_j) + \gamma_7 Type_j + u_j \end{aligned} \tag{3.3}$$

In this model, we use various product and shop characteristics to estimate the average report score. For time-varying variables, we take their values of the first date of observation as an approximate. The unobserved characteristic u_j is correlated with the product random effect ξ_j in Equations (3.1) and (3.2). We also consider the correlation between u_j and the shop's trade-off procedure between providing free samples and boosting future sales. Details on modeling the correlations are in Section 3.2.2. In this way, we can solve the endogeneity issue of the report score.

The correlation between product price and ξ_j is due to the shop's pricing strategy. A shop is more likely to sell a product with unobserved high quality at a high price. In this case, product price and product random effect ξ_j should have a positive correlation. To address the issue of endogenous price, we use instrumental variables for price to eliminate the endogeneity. Details can be found in Section 3.2.3.

3.2.2 *Supply Model*

To solve the problem of endogenous sampling size, we build a model on the supply side describing the decision-making process of shops on how many free samples to offer. A common practice of modeling the decision process for suppliers is to find cost shifters that can affect the marginal cost of the product and then find the relationship between marginal cost and the supplier's best decision (how many free samples to provide, in our case) by solving an optimization problem. In this way, we can estimate the parameters of demand side and supply side simultaneously. This approach is developed and well-explained in the BLP model (Berry et al., 1995). However, as we have only limited information about the product with no information that can perform as cost shifters, we cannot apply this method directly. Instead, we need to find an alternative way to describe the decision process of shops without considering the marginal cost.

Specifically, we introduce a trade-off factor for the shop. A shop would expect that a sampling campaign would attract consumers and boost sales in the future. Additionally, a larger sampling size would attract more consumers. However, sampling more also means losing more money at present, because the shop is sending free products to consumers without receiving revenue from them. Hence, the shop needs to make a trade-off between present loss and future gain. We model the decision process of the shop as follows:

$$\max \Pi_j = \sum_{t=1}^{T} \mathbb{E}\left(Sales_{jt} | Size_j\right) - r_j \times Size_j$$
$$s.t.\ Price_j \times Size_j \geq C \tag{3.4}$$

Here Π_j is the objective function that the shop needs to maximize. $Size_j$ is the sampling size of the product. $\mathbb{E}\left(Sales_{jt} | Size_j\right)$ is the expected sales on the t-th day after the sampling campaign begins (we still use t here to simplify notations). r_j is the trade-off factor. The shop makes the decision by first calculating the expected sales in the period when the effect of sampling is valid and then making a trade-off between expected sales and sampling size under the restriction of sampling threshold C, which is ¥1,500 according to the sampling center's rule. We assume $T = 30$, which means that the effect of the sampling campaign lasts for about a month. We also explore situations where T varies in our robustness check.

We estimate the expected sales of the product using Equations (3.1)-(3.3). Besides $HasReport_{jt}$, $AvgReportScore_{jt}$, and $DaysElapsed_{jt}$, we use the value of variables on the first day of our observations of the product. This is reasonable, as when the shop makes the expectation of future sales, it uses its current characteristics as an approximation of future characteristics. $DaysElapsed_{jt}$ just takes the value of t (i.e., the number of days since sampling began). For variables related to reports, we need to consider report scores and the arrival time of reports. To simplify the calculation of expected sales, we assume that the shop considers the arrival date of only the first report (hereafter the first arrival date) and the average report score. The average report score can be found in our data for sampling products, and it can be estimated by Equation (3.3) for non-sampling products. The first arrival date is assumed to follow a nonparametric distribution which can be derived from our dataset. We can then calculate the expected sales by integrating out first arrival date. Finally, as shops may form biased estimations on the effect of reports (such as overestimating the importance of having feedback reports), we also add $ReportDeviation$ to capture this bias when reports are available.

In Equation (3.4), we should notice that to make the best sampling size solved from this optimization problem positive for some shops, the effect of sampling size on sales has to be positive. As we want to study the impact of sampling size on sales, it may be a too strong assumption that it is positive. However, we find model-free evidence of the positive effect of sampling size in our data. Out of 131 shops participating in the sampling campaign, 39 are unbounded by the binding condition in Equation (3.4). In other words, they sample more than the product quantity required to launch their products in the sampling campaign. As we assume that shops are rational decision makers, these shops cannot offer more products than the binding amount without expecting higher future sales due to a larger sampling size. Thus, our assumption of a positive effect of sampling size is reasonable. Our goal is to find the significance level and the magnitude of this effect. If the magnitude of this effect is too small, it is not an economically significant effect, despite its statistical significance.

The trade-off factor in Equation (3.4) can be inferred by some shop and product characteristics as follows:

$$
\begin{gathered}
ln(r_j) = \alpha_0 + \alpha_1 WorkHour_j + \alpha_2 WorkerNum_j + \alpha_3 Delivery_j \\
+\alpha_4 PromiseNum_j + \alpha_5 Type_j + \omega_j \\
\begin{pmatrix} \xi_j \\ \omega_j \\ u_j \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Omega})
\end{gathered}
\tag{3.5}
$$

Variables in Equation (3.5) are the same as those in Table 3.2. Due to similar reasons as we explain for variables used in Equation (3.4), $Delivery_j$ and $PromiseNum_j$ take values of the first day that we observe from $Delivery_{jt}$ and $PromiseNum_{jt}$. ω_j is the unobserved characteristic of the product and the shop that can affect the trade-off factor. As pairwise correlations may exist in the set of three unobserved characteristics $\{\xi_j, \omega_j, u_j\}$, we model the joint distribution of three random effects to be a multivariate normal distribution with mean zero and variance-covariance matrix $\mathbf{\Omega}$.

This model can capture two channels described in Section 3.2.1 that make sampling size an endogenous variable. We model the decision process of the shop in Equation (3.4) and address the potential correlations between sampling size and the unobserved product quality. We also model the pairwise correlations in $\{\xi_j, \omega_j, u_j\}$. In this way, we can estimate the demand side and supply side simultaneously and address the endogeneity issue of sampling size.

As the shop solves the optimization problem in Equation (3.4) to find its optimal sampling sizes, we can derive the restrictions on the trade-off factor r_j using the observed sampling sizes of the shop. Before finding restrictions for r_j, we need to divide products in our data set into three groups: sampling and bounded, sampling and unbounded, and non-sampling. Sampling and bounded products are those products that take part in sampling campaigns with sampling sizes bounded by the sampling threshold. Sampling and unbounded products are those products that join sampling campaigns with sampling sizes unbounded by the sampling threshold. Non-sampling products are products not taking part in the sampling campaign. Restrictions on the trade-off factor r_j are different for different groups.

When the shop is sampling and bounded, to observe the sampling size that we have in our data set, the restrictions on r_j should make sure that the shop does not benefit from providing one more free sample, and the shop is also not better off if it does not participate in the sampling campaign, namely

$$\sum_{t=1}^{T}\Big(\mathbb{E}\big(Sales_{jt}|Size_j+1\big)-\mathbb{E}\big(Sales_{jt}|Size_j\big)\Big)\leq r_j \tag{3.6}$$

$$\sum_{t=1}^{T}\Big(\mathbb{E}\big(Sales_{jt}|Size_j\big)-\mathbb{E}\big(Sales_{jt}|0\big)\Big)\geq r_j\times Size_j \tag{3.7}$$

When the shop is sampling and unbounded, the sampling threshold exerts no effect on the shop's decision. In this case, the shop's best sampling size without restriction is the sampling size we observe in our data set, which means that the shop cannot be better off if it samples one more or one less product. Thus, the restrictions on r_j are

$$\sum_{t=1}^{T}\Big(\mathbb{E}\big(Sales_{jt}|Size_j+1\big)-\mathbb{E}\big(Sales_{jt}|Size_j\big)\Big)\leq r_j \tag{3.8}$$

$$\sum_{t=1}^{T}\Big(\mathbb{E}\big(Sales_{jt}|Size_j\big)-\mathbb{E}\big(Sales_{jt}|Size_j-1\big)\Big)\geq r_j \tag{3.9}$$

When the shop is non-sampling, it must be the case that the shop finds the sampling threshold too high. The shop is better not participating in the sampling campaign than participating in the campaign with a minimum sampling size. We denote $\underline{Size}_j$ as the minimum sampling size of the product qualified for the sampling campaign. Then the restriction on r_j is

$$\sum_{t=1}^{T}\Big(\mathbb{E}\big(Sales_{jt}|\underline{Size}_j\big)-\mathbb{E}\big(Sales_{jt}|0\big)\Big)\leq r_j\times \underline{Size}_j \tag{3.10}$$

In our supply model, we consider the decision process of the shop as a trade-off between present loss and future gains. We model unobserved characteristics of product quality, trade-off factor, and report score to follow a multivariate normal distribution in Equation (3.5). By jointly estimating demand and supply sides, we can address the endogeneity issues.

3.2.3 *Price Endogeneity*

Price is another potential endogenous variable mentioned in Section 3.2.1. High-quality products usually have high prices, and we cannot observe a product's actual quality. Though in our context, the market on Taobao is very competitive and shops have small market powers to change price, implying that the problem of endogenous price is not severe, we still use instrumental variables to eliminate potential endogeneity in price. We do not model the pricing procedure explicitly, as we have no information on cost shifters and hence cannot estimate the marginal cost of the product.

As we crawled products in pairs (one sampling product and one non-sampling product) and every product is very similar to its paired product except in product quality, we use the price of the paired product as an instrumental variable (IV) for the focal product. The price of the paired product is not correlated with the quality of the focal product, and it provides enough information about the exogenous part of the price, so it should be a strong instrument. Moreover, as shown in our subsequent results, the paired price indeed serves as a good instrumental variable.

3.2.4 *Estimation*

We develop a two-stage estimation procedure to estimate our structural model. In the first stage, we regress price on paired price and other exogenous variables (excluding sampling size and report variables) to estimate exogenous price. We replace price with the exogenous price to solve the problem of price endogeneity.

In the second stage, we jointly estimate parameters in Equations (3.1), (3.3), and (3.5) (for the homogeneous case) or Equations (3.2), (3.3), and (3.5) (for the heterogeneous case). We employ the simulated maximum likelihood estimation (Simulated MLE) to estimate parameters. When forming the likelihood of a product (or a shop), we need to consider the likelihood of observing sales and report scores in data and the probability that the trade-off factor r_j is in the interval characterized in Section 3.2.2. Finally, the likelihood functions for sampling shops (Equation (3.11)) and non-sampling shops (Equation (3.12)) can be calculated as follows:

$$L_j^S(\alpha,\beta,\gamma,\sigma_\varepsilon,\mathbf{\Omega}) = f_u(u_j;\gamma,\mathbf{\Omega})\int_{d_j} P(\xi_j,\omega_j|u_j,d_j;\alpha,\beta,\sigma_\varepsilon,\mathbf{\Omega})dF_d(d_j) \tag{3.11}$$

$$L_j^{NS}(\alpha,\beta,\gamma,\sigma_\varepsilon,\mathbf{\Omega}) = \int_{d_j}\left(\int_{u_j} P(\xi_j,\omega_j|u_j,d_j;\alpha,\beta,\sigma_\varepsilon,\mathbf{\Omega})f_u(u_j;\gamma,\mathbf{\Omega})du_j\right)dF_d(d_j) \tag{3.12}$$

where $P(\xi_j,\omega_j|u_j,d_j;\alpha,\beta,\sigma_\varepsilon,\mathbf{\Omega})$ can be calculated by:

$$P(\xi_j,\omega_j|u_j,d_j;\alpha,\beta,\sigma_\varepsilon,\mathbf{\Omega}) = \int_{\xi_j} L^D(\xi_j;\beta,\sigma_\varepsilon)\left(\int_{\underline{\omega}_j(\alpha,\beta,\sigma_\varepsilon)}^{\overline{\omega}_j(\alpha,\beta,\sigma_\varepsilon)} f_{\omega|\xi,u}(\omega_j|\xi_j,u_j;\mathbf{\Omega})d\omega_j\right)f_{\xi|u}(\xi_j|u_j;\mathbf{\Omega})d\xi_j \tag{3.13}$$

Here σ_ε is the standard deviation of the idiosyncratic error in Equation (3.1) or (3.2), and d_j stands for the first arrival date. We apply nonparametric distribution for d_j. We can calculate the probability of observing report scores by $f_u(u_j;\gamma,\mathbf{\Omega})$ for sampling shops, while we do not know report scores and need to integrate out the probability of u_j for non-sampling shops. When d_j and u_j are fixed, we can calculate the conditional probability of (ξ_j,ω_j) by Equation (3.13). The likelihood of observed sales is calculated by $L_D(\xi_j;\alpha,\beta,\sigma_\varepsilon)$ for a fixed ξ_j. We can then calculate the lower and the upper bounds of ω_j ($\underline{\omega}_j(\alpha,\beta,\sigma_\varepsilon)$ and $\overline{\omega}_j(\alpha,\beta,\sigma_\varepsilon)$) by solving ω_j out from restrictions on r_j. As ξ_j and ω_j follow a bivariate normal distribution conditional on u_j, we can calculate the likelihood of the supply side by conditional likelihood $f_{\omega|\xi,u}(\omega_j|\xi_j,u_j;\mathbf{\Omega})$. Finally, we can integrate the likelihood under fixed ξ_j over ξ_j to get the likelihood of each product. To simplify the expression of likelihood, we integrate out the joint likelihood of sales and supply side conditional on u_j. Details can be found in Appendix A.1. Finally, we simulate values of d_j and u_j in Equations (3.11) and (3.12) to find the simulated likelihood and maximize it on the whole dataset.

3.3 RESULTS

3.3.1 *Preliminary analysis*

We first conduct a preliminary analysis by estimating random effect models in Equations (3.1) and (3.2) without considering the endogeneity of sampling size, price, and report score. The results are in columns (1) and (2) of Table 3.3. From column (1), we find that the effect of sampling size is positive and significant, but the effect of price is positive and insignificant. As we have not used instrumental variables to control the endogenous price, the coefficient of price is biased upwards and becomes insignificant. Column (2) shows the heterogeneous effect of sampling size. We find that experience goods benefit the most from the increase of sampling size. However, as the endogeneity of sampling size is not eliminated, the moderating effects can also be biased.

Though our preliminary analysis does not consider the problem of endogeneity, it can still provide some evidence on the positive effect of sampling size on sales and the potential heterogeneity of the effect. To solve the problem of endogenous sampling size, price, and report and show the mechanism of shops' decision on sampling size at the same time, we need to estimate the structural model we built in the previous section.

3.3.2 *Results of Structural Model: Homogeneous Case*

We use a two-stage estimation for the structural model. In the first stage, we regress price on instrumental variables and exogenous variables. The R^2 of this regression is 0.8935, and it has an F statistic much higher than 10, which indicates that paired price is a strong IV for focal price. We use the paired price as the IV to estimate the exogenous price. We replace the original price with the estimated exogenous price in our second stage to address the endogeneity issue of price. Details of the results can be found in Appendix A.2.

In the second stage, we estimate the structural model built in Sections 3.2.1 and 3.2.2, and we present the results of homogeneous and heterogeneous cases in columns (3) and (4) of Table 3.3, respectively. The effect of price is negative and significant in both results after using instrumental variables to control the endogenous price. The results in column (3) show that sampling size has a significant effect on sales quantity. A 1% increase in sampling size can raise daily sales quantity by 0.178%. Though this daily effect is not tremendous, the effect lasts for a period (such as a month, in our assumption), which can provide shops considerable return. As we set $T = 30$ in our model, we can estimate that a 1% increase in sampling size can bring the shop a total sales increase of 5.34% daily sales over a month. Thus, the effect of sampling size on sales is both statistically and economically significant. This finding provides evidence of the signaling effect in physical goods sampling via online channels, which resembles the promotional effects in other sampling schemes (Bell et al. 2015; Gallino and Moreno 2018; Wei et al. 2021). It also adds to the literature on similar signaling effects in informational goods sampling (Wang and Özkan-Seely 2018). Because a shop with higher product quality would expect more future sales and sample more, we find that a positive correlation between product quality and sampling size is built through the first channel that makes sampling size an endogenous variable. This concurs with customers' opinion about the sampling campaign that a high-quality product would sample more to gain more future sales. And it also verifies how the signaling effect of sampling size works in our context.

However, to fully solve the endogeneity issue of sampling size, we also need to find the correlation between product quality and the trade-off factor, the second channel of endogeneity mentioned in Section 3.2.1. We observe a strong positive correlation between shops' daily sales quantity and their trade-off factor r_j, which is denoted as $\rho_{\xi\omega}$ in Table 3.3. This indicates that a shop selling a product of higher quality is more likely to give a higher weight to the loss from free samples when it makes the trade-off between present loss and future gain. This is an interesting phenomenon explained by the fact that shops with higher product quality are confident about their products and show less interest in sampling campaigns when considering only the second channel of endogeneity. On the contrary, shops with lower quality would rush to participate in sampling campaigns because they are afraid that the quality of their products cannot naturally attract enough customers to make profits without taking part in sampling campaigns. The consequence of the above behaviors of shops is that shops with higher product quality offer fewer free samples than shops with lower product quality when their expected sales are the same. This channel presents a negative correlation between product quality and sampling size, which is opposite to the direction of correlation through the first channel. Thus, the correlation between sampling size and product quality is complicated. However, the positive and significant effect of sampling size supports that customers still think the overall correlation between sampling size and product quality is positive, and sampling more signals better quality.

We then focus on the effect of feedback reports on sales and endogeneity issues related to reports. According to our results, the feedback report has a significant impact on sales. An average report score of 5 has a positive effect on sales, yet an average report score lower than 4.5 hurts shops. Interestingly, we find that shops usually overestimate the effect of having feedback reports for their sampling products. The results also show that the effect of price on report score is negative and significant, which justifies adding the interaction term of ln $(Sales)$ and $Sample$ in the sales equation to alleviate the endogeneity problem of price. Another observation from the results is that the correlations between the random effect of the average report score and the other two random effects are small and insignificant. This means that the unobserved characteristic of the report score is relatively independent, and the feedback report does not have severe endogeneity issues.

Table 3.3. Effect of Sampling Size and Moderating Effects

	(1) Random Effect	(2) Random Effect +Moderating	(3) Structural+IV	(4) Structural+IV +Moderating
Demand Equation:				
ln(Size)	0.146*** (0.024)	0.191*** (0.037)	0.178*** (0.022)	0.239*** (0.034)
*ln(Size) * Search*	/	-0.056 (0.047)	/	-0.093** (0.037)
*ln(Size) * Credence*	/	-0.092* (0.055)	/	-0.072* (0.042)
ln(Price)	0.025 (0.037)	0.023 (0.037)	-0.255*** (0.036)	-0.261*** (0.036)
*ln(Price) * Sample*	0.033 (0.021)	0.036* (0.021)	0.199*** (0.019)	0.203*** (0.019)
HasReport	-2.383*** (0.609)	-2.392*** (0.610)	-2.003*** (0.478)	-1.856*** (0.443)
ReportScore	0.516*** (0.124)	0.517*** (0.124)	0.441*** (0.098)	0.409*** (0.092)
DaysElapsed	-0.024*** (0.003)	-0.024*** (0.003)	-0.024*** (0.003)	-0.024*** (0.003)
Delivery	0.104 (0.109)	0.111 (0.109)	0.097 (0.114)	0.105 (0.115)
PromiseNum	-0.063** (0.030)	-0.065** (0.030)	-0.071** (0.032)	-0.073** (0.032)
ln(ReviewVolume)	0.168*** (0.018)	0.165*** (0.018)	0.101*** (0.016)	0.101*** (0.016)
ShopRating	0.158 (0.447)	0.163 (0.449)	0.119 (0.395)	0.139 (0.343)
ln(ShopRatingVolume)	0.060*** (0.022)	0.060*** (0.022)	0.041** (0.017)	0.038** (0.017)
Search	0.034 (0.137)	0.109 (0.149)	0.383** (0.163)	0.502*** (0.170)
Credence	-0.035 (0.156)	0.078 (0.170)	0.199 (0.187)	0.296 (0.195)
DayTypeDummies	√	√	√	√
Constant	-1.118 (2.191)	-1.190 (2.197)	0.152 (1.936)	0.013 (1.673)
σ_ξ	0.736*** (0.035)	0.739*** (0.035)	0.902*** (0.050)	0.906*** (0.050)
σ_ε	0.836*** (0.007)	0.836*** (0.007)	0.835*** (0.007)	0.835*** (0.007)
Report Equation:				
ln(Price)	/	/	-0.048** (0.022)	-0.047** (0.022)
Delivery	/	/	-0.094 (0.078)	-0.093 (0.078)
PromiseNum	/	/	0.032* (0.017)	0.032* (0.017)
ln(ReviewVolume)	/	/	0.013 (0.010)	0.013 (0.010)
ShopRating	/	/	-0.040 (0.310)	-0.044 (0.309)
ln(ShopRatingVolume)	/	/	-0.018 (0.011)	-0.018 (0.011)
Search	/	/	0.084 (0.062)	0.081 (0.061)
Credence	/	/	-0.042 (0.068)	-0.044 (0.068)
Constant	/	/	5.232*** (1.506)	5.245*** (1.502)
σ_u	/	/	0.218*** (0.014)	0.217*** (0.014)
$\rho_{\xi u}$	/	/	-0.003 (0.108)	-0.005 (0.110)
Supply Equation:				
WorkHour	/	/	0.002 (0.002)	0.002 (0.002)
WorkerNum	/	/	-0.009 (0.028)	-0.010 (0.028)
Delivery	/	/	-0.226 (0.260)	-0.173 (0.262)
PromiseNum	/	/	-0.045 (0.066)	-0.041 (0.067)
Search	/	/	1.467*** (0.267)	0.971*** (0.331)
Credence	/	/	0.642** (0.309)	0.242 (0.391)
Constant	/	/	1.603*** (0.454)	1.917*** (0.457)
ReportDeviation	/	/	0.941*** (0.171)	0.911*** (0.161)
σ_ω	/	/	1.410*** (0.091)	1.421*** (0.093)
$\rho_{\xi\omega}$	/	/	0.886*** (0.024)	0.890*** (0.023)
$\rho_{\omega u}$	/	/	0.112 (0.124)	0.103 (0.126)
Number of observations	7,076	7,076	7,076	7,076
AIC	/	/	19,022.01	19,019.50

*Note: Standard errors in parentheses; Significant levels: *** p <0.01, ** p <0.05, * p <0.1*

3.3.3 *Results of Structural Model: Heterogeneous Case*

In column (4), we show the moderating effect of product type on the impact of sampling size. We find that search products and credence products benefit significantly less than experience products from increasing sampling size, consistent with the preliminary analysis. Unlike the preliminary analysis, which does not control endogenous sampling size and price, our structural model shows that search products are affected least by sampling size. The reason is that the quality of search products can be easily inferred from the information displayed on the product website. Hence, customers are not significantly affected by sampling —they do not need free samples to know the quality of the product. Credence products are also affected less than experience goods by sampling size. This seems counter-intuitive, as the quality of credence goods is the hardest to infer without sampling campaigns, and customers should be affected the most by their sampling size. However, if we recall that the quality of credence goods is hard to infer even after usage, it is reasonable that the effect of sampling size on credence goods is mediocre. Because free samples still cannot show the quality of credence goods, customers would not think that offering more free samples signals better quality. Thus, compared to experience goods—whose quality is hard to infer before usage but easy to infer after usage—credence goods benefit less from increasing sampling size. But they still benefit more than search goods, according to our results. The results of our heterogeneous model add to the literature on emerging online promotional strategies that resemble product sampling (Wei et al. 2021) and the signaling effect literature (Lee et al. 2018; Wang et al. 2019) that investigate heterogeneous effects.

3.3.4 *Policy Counterfactuals on Sampling Threshold*

As Taobao sets a sampling threshold for shops interested in sampling, we are interested in what effect this sampling threshold exerts on shops' decisions in the sampling campaign. Using estimated parameters in our results, we can simulate counterfactual scenarios where the sampling threshold changes. We investigate how this policy change affects the number of sampling shops and average sales for all shops.

We simulate both the case where the sampling threshold increases and the case where it decreases. That is, we allow the sampling threshold to vary from ¥200 to ¥3,000. We consider the decisions of all shops in our data. We focus on the number of sampling shops, which describes the actual size of the sampling campaign, and average sales quantity for all shops, which reflects the performance of e-tailers. The results of our simulation are presented in Figure 3.3.

The number of sampling shops decreases with the sampling threshold, and the decreasing rate slows down when the sampling threshold rises. The logic behind the decreasing pattern is intuitive, as a lower threshold attracts more shops that cannot afford to offer free samples when the threshold is high. The interesting phenomenon here is that the decreasing rate is decreasing. It indicates that reducing the threshold can expand the actual size of the sampling campaign by a large amount, whereas raising the threshold shrinks the actual size of the sampling campaign by only a small amount. We can also derive some quantitative results here. Compared to the current threshold, a sampling threshold of ¥200 can attract 57.2% more shops, and a sampling threshold of ¥3,000 closes the door to sampling campaigns for 21.2% sampling shops. E-commerce platforms can leverage these results to better control the size of their sampling campaigns.

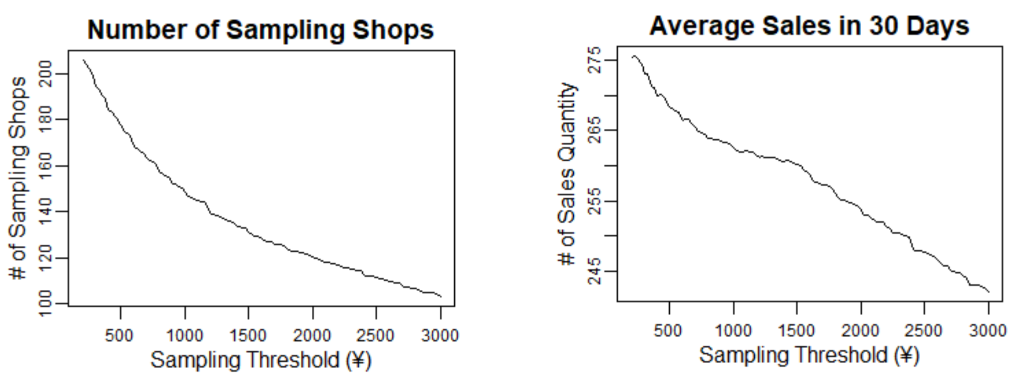

Figure 3.3. Policy Simulation Results

The average sales quantity of all shops is also decreasing in the sampling threshold, yet its pattern is slightly different, and the mechanism is more complicated. The decreasing rate of average sales does not change much, especially when the sampling threshold is large. The monotonicity of average sales can be explained as follows. When the sampling threshold increases, some shops decide to still sample and sample more, whereas some shops decide to leave the campaign. The remaining sampling shops have larger sampling sizes, which results in higher sales. However, exiting sampling shops suddenly become non-sampling shops, and they lose much more sales as the marginal effect of sampling size is diminishing. As a result, average sales quantity monotonically decreases with the sampling threshold. As for quantitative analysis, reducing the sampling threshold to ¥200 can boost average sales by 5.83%, and increasing the sampling threshold to ¥3,000 will lower average sales by 6.99%. Our result provides the consequences of varying the sampling threshold on average sales and can offer advice for e-commerce platforms when choosing the sampling threshold.

3.3.5 *Robustness Checks*

We perform several robustness checks in this section. In section 3.2.2, we assume that the shop expects the effect of sampling to last for 30 days (i.e., $T = 30$). However, the shop may have a different expectation about the duration of the sampling effect. To check if our main results hold when the expected duration of the sampling effect changes, we explore homogeneous cases where the shop expects the sampling effect to last 20, 40, and 60 days. The table presenting our results can be found in Appendix A.3. Our results indicate that the impact of sampling size is robust with respect to the expected duration of the sampling effect. We also notice that the AIC (Akaike information criterion) of the model decreases with the expected duration. As the AIC of our main result is very close to that of the model with 40 or 60 days of expected duration, we regard 30 days as a reasonable assumption of the sampling effect's duration in our structural model.

We also check the robustness of our results regarding monthly sales, campaign status, and time. The results can be found in Appendix A.3. As consumers can observe monthly product sales, we control it to further alleviate the endogeneity issue. The results show that monthly sales has a small and significant effect on sales. Because we have already used random effects to control unobservable product characteristics, adding monthly sales does not change our main result. From another aspect, whether the sampling campaign is ongoing or completed can moderate the effect of sampling size. The results of this moderating model show that our main result holds and that the signaling effect of sampling size is stronger when the sampling campaign is completed. We further check if the impact of sampling size lasts for a long time using a relative time effect model, and the results indicate that it is indeed a long-term effect.

3.3.6 *Discussion*

The endogeneity issues of sampling size, price, and feedback reports are all addressed in the structural model based on our results. Without controlling for price endogeneity using instrumental variables, the price coefficient is biased upward and becomes positive and insignificant. The positive correlation between price and unobserved quality contributes to the endogeneity of price. However, using the structural model seems to have little effect on estimating the coefficient of sampling size. This phenomenon is an integrated effect of two channels that make sampling size correlated with unobserved product (or shop) quality, which is explained in Section 3.2.1 in detail. According to our results, on the one hand, a product of higher quality would expect more future sales and thus sample more. On the other hand, shops with high-quality products do not care much about sampling campaigns and put less weight on more future sales resulting from the campaign, so they tend to sample less. The negative correlation from the second channel neutralizes the positive correlation from the first channel. The integration of two channels of correlation gives us the impression that there is little correlation between sampling size and unobserved product quality. Using our structural model, we can determine the actual mechanism related to sampling size and get an unbiased estimate of the effect of sampling size.

The results of the homogeneous structural model show that the positive effect of sampling size is both statistically significant and economically significant. This proves the signaling role that sampling size plays in physical goods sampling on online platforms and helps e-tailers and online platforms to understand the effect of sampling size. Our results shed light on the mechanism behind physical goods sampling on online platforms, which is very different from the mechanisms of offline physical goods sampling and online information goods sampling.

Our findings on the moderating effect of product types can also provide important managerial implications for shops seeking to leverage sampling campaigns to boost their sales. Search goods benefit the least from sampling campaigns, as expected, as their quality can be inferred before using them. Shops selling search goods should be well aware of this fact and not expect much return from offering free samples. What goes beyond our expectation is that credence goods benefit less than experience goods from sampling campaigns. Customers are not confident enough about the quality of credence goods even after using free samples because they cannot discover the true quality. This finding advises shops selling credence goods that their sampling campaigns may not benefit their sales as much as they expect. Observing our results, they can have reasonable expectations for the effects of their sampling campaigns.

3.4 CONCLUSIONS

Free product sampling acts as an effective promotional method in the traditional market. With the development of the Internet, it has also been adopted as a popular marketing strategy in many online markets. Recently, physical goods sampling via online channels has emerged in many e-commerce platforms. It works differently from offline physical goods sampling and online information goods sampling because it serves as a signal sent from shops to customers. However, compared to the other two sampling schemes, this new form of sampling has received little attention in academia. In this paper, we fill this research gap by investigating the effect of sampling size (i.e., the quantity of sampling product) on sales as well as how product type (search, experience, and credence) moderates this effect. We also derive some results on what role the sampling threshold plays in the sampling campaign.

Leveraging a data set from Taobao, the largest e-commerce platform in China, we construct a structural model to estimate the effect of sampling size on sales. To address the endogeneity issue of sampling size, we adopt the idea of BLP to characterize shops' decision process on how much to sample. We find that sampling size has a positive and significant effect on future sales. Experience goods benefit the most from sampling campaigns, and search goods benefit the least. Shops selling credence goods may find this result most helpful, as they may overestimate the effect of sampling on their sales. Utilizing our structural model, we can simulate situations where the sampling threshold changes. We find that the number of sampling shops and average sales of shops both decrease with the sampling threshold. We also provide quantitative analysis on the impact of the sampling threshold to help platform managers set the appropriate sampling threshold.

Our study contributes to the literature in several aspects. First, we enrich the body of studies on product sampling. The nature of online platforms makes the mechanism behind sampling campaigns on these platforms very different from that of traditional offline contexts (Bell et al. 2015; Gallino and Moreno 2018; Wei et al. 2021) and online information goods contexts (Cheng et al. 2015; Cheng and Tang 2010; Wang and Özkan-Seely 2018). Leveraging a BLP-style structural model, we provide empirical evidence on the signaling effect of sampling size, which is a unique mechanism in physical goods sampling via online channels. Second, our work adds to the literature on the signaling effect on e-commerce platforms. Previous studies have documented various signals provided by consumers, platforms, and sellers (Lee et al. 2018; Wang et al. 2019). Our research identifies sampling size as another valid signal for e-tailers (i.e., sellers) to show their product quality. Third, we contribute to the literature of product types by investigating the moderating effect of product types on the signaling effect of the sampling size. In contrast to previous studies stating that only experience goods participate in sampling campaigns, our study finds that search goods and credence goods also benefit from sampling campaigns held online, though they benefit less than experience goods. Finally, we construct a novel BLP-style structural model capturing the decision process of e-tailers to address the endogeneity issues in our context. Our model can be further used in studies of online platforms where endogeneity is a severe problem and people's decision processes cannot be neglected.

Our work can potentially be extended in several ways. First, we have no access to the detailed information of customers. Future studies can refine our model by building a discrete choice model for customers to explain their behavior and make more policy simulations possible if more information about customers is available. Another direction of extension is to estimate the duration of the sampling effect. As we collect data of shops only in a month, we cannot investigate how many days the sampling effect can last. Future studies may also examine the impact of related signals, such as application counts of free samples on online platforms. We hope our study can provide enough insights for future research investigating various phenomena related to physical goods sampling via online channels, especially valuable signals that take advantage of online platforms' unique context.

Chapter 4. WHEN TO BROADCAST? INVENTORY DISCLOSURE POLICIES FOR ONLINE SALES OF LIMITED INVENTORY

Online retail sales have gained popularity since the last decade, especially with the COVID pandemic dramatically accelerating the shift of sales from the brick-and-mortar stores to online retail platforms. Consumers in the United States were estimated to spend $709.78 billion from online retail platforms or, equivalently, 14.5% of the total retail sales in 2020, “representing both an all-time high and the biggest share increase in a single year” (Lipsman and Liu 2020).

Many online sales are offered with limited inventory or for a short sale window, so a sale ends either once all inventory is gone or the sale period ends. In particular, our study is motivatedby online flash sales (*a.k.a.* lightning or daily deals), which are essentially online sales of limited inventory *and* within a short period of time. Such a format of sales has gained popularity not only among e-tailers that focus solely on flash sales, such as Woot.com, Groupon, Zulily, but also among some leading online platforms, such as Amazon lightning deals and eBay daily deals. In recent years, a number of studies in the operations management literature have proposed various ideasfor improvement in a flash sale platform’s demand forecasting techniques and pricing strategies. For instance, Ferreira et al. (2016) and (2018) advocated a dynamic pricing policy for a flash saleplatform, which, however, has not been embraced by the majority of the flash sale platforms so far, as most of the platforms are still offering a fixed discount for a product during a relatively short sale period. On the other hand, some studies proposed leveraging the platform’s inventory information provision. Notably, Calvo et al. (2020) and Cui et al. (2019) explored the impact of inventory availability information on demand in online flash sales. Cui et al. (2019) found that“a decrease in product availability causally attracts more sales in the future; in particular, a 10% increase in past claims leads to a 2.08% increase in cart add-ins in the next hour."

Inspired by the empirical studies concerning the impact of inventory information on demand inthe context of online sales with limited inventory (e.g., flash sales), we develop an analytical modelto optimize a platform's inventory disclosure policy. The question is: what is the best timing for the platform to reveal the remaining inventory? To the best of our knowledge, our study is the first analytical work that addresses this question for online retail platforms. Various inventory disclosure policies have been employed in practice. Some sales do not broadcast inventory availability, which we refer to as the "*never disclose*" policy, whereas some other sales, including Amazon's lightning deals, reveal inventory levels from the beginning of the sales, which we refer to as the "*always disclose*" policy. A more sophisticated disclosure policy employed in practice is: wait until the inventory level drops below a threshold to disclose availability of the product, which we refer to as a *fixed threshold* policy. In their study which was based on data from a multinational retail platform, Calvo et al. (2020) noted that the retail platform employed such a disclosure policy with a fixed inventory threshold. Another example is eBay, which reveals the remaining inventory onlywhen the inventory level becomes sufficiently low.

In addition to the aforementioned common inventory disclosure policies, we propose and analyze an improvement on the fixed threshold policy, which we refer to as a *time-dependent threshold* policy. Note that, in the fixed threshold policy, the inventory level is revealed once it drops to a pre-determined threshold, no matter when the threshold is reached. There is, however, clearly room for improvement in such a policy as it does not consider the precise time when the inventory level reaches the fixed threshold. Indeed, our proposed policy indicates that the platform's inventory disclosure decision should be based on not only the inventory level but also the remaining time of the sale period. The intuition behind the proposed policy is that the firm should disclose its inventory only if the inventory level drops faster than expected, thus sending the customers a clearsignal that the deal is of high quality and the product is a popular item.

In summary, in this paper we examine four inventory disclosure policies: (i) the *"never disclose"* policy, (ii) the *"always disclose"* policy, (iii) the *fixed threshold* policy, and (iv) the proposed *time- dependent threshold* policy. We would like to point out that, although the two simple policies are widely employed in practice, optimization of the fixed and time-dependent threshold policies, which is the focus of this study, can significantly improve the baseline performance of those simple policies. Below, we summarize the contributions of this study:

- Although the existing empirical studies have established the impacts of inventory information disclosure on customers' purchasing behavior, they have not examined analytically how an online retail platform can improve its inventory disclosure policy leveraging such salient impacts. To the best of our knowledge, our study is the first to fill in the gap between the empirical findings and optimization of the platform's inventory disclosure policies.
- We develop an analytical model to examine the aforementioned inventory disclosure policies. Specifically, under the fixed threshold policy, we derive the probabilistic distribution of the future sales once the platform reveals its inventory information, which plays a key role in formulating the expected total sales based on any given threshold. Then, the optimal threshold can be found efficiently. For a flash sales platform which currently employs a fixed threshold policy, our algorithm can be easily implemented to find the

optimal threshold and thus achieve the best performance of that policy. Further, we show that the time-dependent threshold policy is the optimal policy under two reasonable assumptions, and we devise an efficient dynamic-programming-based algorithm to find the optimal parameters of the optimal policy.

- Through a comprehensive numerical study, we examine the performances of different policies and the associated optimal thresholds. To this end, we provide a customer choice model within a Bayesian updating framework, which captures two important effects of the inventory information on customer demand: the observational learning (a.k.a. herding) and scarcity effects. Then, we derive the functional form of the average demand rates through aggregating individual customer's purchasing behavior. We find that both threshold policies significantly outperform the two simple policies, and that the fixed threshold policy is near optimal in cases where the herding effect dominates the scarcity effect. However, the threshold policies may backfire when the customers perceive not disclosing as an important signal of slow sales.

In summary, our study provides an efficient analytical solution for policy optimization as well as guidelines for policy selection in light of various demand characteristics. It is worth noting that our model and solution procedures can find broad applications, including but not limited to flash sales, as long as the following two conditions hold: (1) The supply of the product is not abundant or the planning horizon only lasts for a limited period of time (or both); (2) information disclosure enables the customers to infer the value or scarcity of the product, thus disclosure of past sales and limited inventory stimulating the future demand. For example, on its common sales other than the lightning deals, Amazon does not provide inventory information when there is still plenty of inventory (e.g., "in stock") but discloses inventory information once the inventory level is low (e.g., "only 3 left in stock – order soon"), which resembles the fixed threshold policy examined in this paper. Although the common sales on Amazon typically last longer than its lightning deals (flash sales), customer purchasing decisions will be influenced by the disclosure of inventory information, so the solution and insight of our model can possibly apply.

4.1 Model and Analysis

Consider a sale that starts with K units of inventory and lasts T time periods. We assume that the arrivals of potential customers for the product follow a non-homogeneous Poisson process. For example, upon arrival, a customer who browse the website may infer the value of the deal from actions of other customers or the probability that the product will be sold out, which depends on the availability of inventory information. In this section, we omit details about how the average demand rate can be derived from individual customers' purchasing propensity, which we defer to Section 4.2. Here, we focus on analyzing and optimizing the platform's inventory disclosure policies based on a general form of the demand functions. Overall, we need to distinguish the average demand rates in three different scenarios:

1. Customers know that the platform employs a "never disclose" policy such that the inventory information will never be disclosed; in such case, the average demand rate can be denoted by $\theta^{ND}(t)$ ($t \in [0, T]$) in general, where "ND" stands for "never disclose."
2. Customers know that the platform employs a threshold policy, but the inventory information is not available at the present time (while it may be disclosed in the future); in such case, the average demand rate is denoted by $\theta^{BD}(t)$, where "BD" stands for "before disclosure."
3. Customers know that the platform employs a threshold policy and has disclosed its inventory information; in such case, there has been empirical evidence (Cui et al. 2019

and Calvo et al. 2020) which suggests that the average demand rate is influenced by the disclosed inventory levels, which we refer to as the *inventory-dependent demand rate*, denoted by $\lambda_i(t)$ in general, where $i \in \{0,1,\cdots,K\}$ is the amount of inventory sold and $t \in [0,T]$.

One may conjecture $\theta^{BD}(t) \leq \theta^{ND}(t)$. The reasoning behind this assumption could be that, although inventory information is not available in both the “BD” and “ND” scenarios, customers in the “BD" scenario may interpret the platform’s not disclosing at the present time as a signal of slow sales as they know the threshold for disclosure has not been reached, which gives a negative impact on the customers’ propensity of purchasing. We believe this is a reasonable assumption; that said, whether it holds is not critical to the optimization of the inventory-disclosure policies.

We now introduce the analytical model. Without loss of generality, let $\lambda_K(t) = 0$ as the sale will end once all available inventory is sold out. In practice, the platform can broadcast either the absolute amount or the percentage of inventory sold. Let $s(i) = i/K$ denote the fraction of inventory sold and $\tau(t) = t/T$ as the fraction of time elapsed. Thus, $\lambda_i(t)$ can be written as follows:

$$\lambda_i(t) = \beta(i,t)\theta^{BD}(t) \text{ or, equivalently, } \lambda_i(t) = \beta(Ks(i), T\tau(t))\theta^{BD}(t).$$

We refer to $\beta(i,t)$ as the *demand adjustment factor*. This factor can be greater than one if the inventory level drops quickly, and vice versa. As noted in the introduction, Cui et al. (2019) found that “a decrease in product availability causally attracts more sales in the future; in particular, a 10% increase in past claims leads to a 2.08% increase in cart add-ins in the next hour."

Currently, Amazon’s lightning deals broadcast the fraction of inventory sold, namely, “% claimed," throughout the sale period. Such an inventory disclosure policy, however, may not be prudent as the future demand rate may slow down when the inventory level is high at the current time, as the customer may read the high inventory level as a signal that the sale is slow and the deal is not of high value. Thus, the key question is: when should the platform disclose its inventory on the basis of the amount and timing of sales made?

In the subsequent analysis, we first consider an inventory disclosure policy based on solely the amount (or percentage) of inventory sold. That is, disclose the inventory level once the number of sales reaches j (or the inventory level drops to $K - j$), where $j \in \{0,1,\cdots,K\}$. We refer to such a policy as the *fixed threshold* policy, which will be analyzed in Section 4.1.1. Two special cases of such a policy is worth noting: $j = 0$ implies disclosure of inventory from the very beginning of the sale, while $j = K$ implies concealing inventory during the entire sale period. Furthermore, having realized that the fixed threshold policy does not take into account when the threshold is reached, we propose an improvement on the fixed threshold policy, namely, the *time-dependent threshold* policy, to be analyzed in Section 4.1.2. Under this policy, the disclosure decision will be made based on not only the amount but also the timing of sales made. Indeed, with certain assumptions about the functional forms of the average demand rates, the proposed policy is optimal.

For the readers' convenience, Table 4.1 summarizes notations throughout this paper.

Table 4.1. Summary of Notations

Notation	Explanation
K	Number of the total available inventory
T	Length of the sale period
$\theta^{ND}(t)$	Average demand rate if the inventory information will never be disclosed, where $t \in [0,T]$
$\theta^{BD}(t)$	Average demand rate before the inventory information is disclosed, given that customers know the platform employs a threshold policy, where $t \in [0,T]$
$\Theta^{\ell}(a,b)$	Integral of the average demand rate without inventory information from time point a to b; that is, $\int_a^b \theta^{\ell}(t)dt$, where $\ell \in \{ND, BD\}$
$\lambda_i(t)$	Inventory-dependent (demand) rate when inventory levels are broadcast, given that i sales have been made until t, where $i \in \{0,1,\cdots,K\}$ and $t \in [0,T]$; let $\lambda_K(t) = 0$
$\beta(i,t)$	Demand adjustment factor due to the disclosure of inventory information: $\beta(i,t) = \lambda_i(t)/\theta^{BD}(t)$; it is equivalent to $\beta(Ks(i), T\tau(t))$ where $s(i) = i/K$ and $\tau(t) = t/T$
$p(x,\mu)$	Probability mass function of the Poisson distribution with rate μ
$P(x,\mu)$	Probability that the Poisson distribution with rate μ is greater or equal to x

$Q_i(t\|\zeta, j)$	Probability mass function of the event that there will be i additional sales during $[\zeta, t]$, given that inventory level is broadcast and there have been j sale at $\zeta \leq t$
$S^*(j)$	Expected total number of sales under a fixed threshold policy, whereby inventory will be broadcast once the number of sales reaches j ($j \in \{0,1, \cdots, K\}$)
$\Gamma(T\|\zeta, j)$	Expected number of sales during $[\zeta, T]$, given that inventory level is broadcast and there have been j sales until the current time ζ, where $\zeta \leq T$
j^*	Optimal inventory-level threshold for disclosure under the fixed threshold policy; the firm should disclose inventory once the sales reach j^* (or the inventory level drops to $K - j^*$)
t_j^*	Optimal time threshold for disclosure under the time-dependent threshold policy; the firm should disclose its inventory if the time of the j-th sale is earlier than t_j^*
$S^*(T\|t, j)$	Optimal expected "sales to go" during $[t, T]$ under the time-dependent threshold policy, given that there have been j sales and inventory information has not been broadcast until the current time t (but the platform may decide to broadcast it from now on)
$V^*(T\|t, j)$	Expected "sales to go" during $[t, T]$ under the time-dependent threshold policy, given that there have been j sales, inventory information has not been broadcast until the current time t, and the platform decides to keep concealing the information for the time being
$q_i(y\|t, j)$	Probability density function of the following event: given that there have been j sales and the platform has not disclosed its inventory until t, it will disclose inventory at $[y - dy, y]$ when the i-th additional future demand occurs, where $i \in \{1, \cdots, K - j\}$
$r_i(y\|t, j)$	Probability of the following event: given that there have been j sales and the platform has not disclosed its inventory until t, there will be i additional sales while inventory will not be broadcast during $[t, y]$, where $i \in \{1, \cdots, K - j\}$
$a \wedge b$ (or $a \vee b$)	The minimum (maximum, respectively) between any two values a and b

4.1.1 *Fixed Threshold Policy*

Now, we consider the fixed threshold policy as follows: when the number of sales is less than j (an integer between zero and K), no inventory information is broadcast. Once the number of sales reaches j, inventory information will be broadcast henceforward. The goal is to find out the optimal j, denoted by j^*, which maximizes the expected total number of sales during the sale period. We would like to point out that, for the analysis of the fixed threshold policy, there is no need to impose any restriction on the functional forms of the inventory-dependent rates.

We start the analysis by deriving the expected number of sales after the platform discloses its inventory. Suppose there have been j sales until the current time ζ, where $\zeta \in [0, T]$, and the platform decides to broadcast its inventory from now on. We need to derive the expected total number of sales during the remaining sale period, $[\zeta, T]$. To this end, let $Q_i(t|\zeta, j)$ be the probability mass function of having i additional sales during $[\zeta, t]$ for any given $\zeta \leq t \leq T$.

For $i = 0$, we have

$$Q_0(t+\Delta|\zeta,j) = \big(1-\lambda_j(t)\Delta\big)Q_0(t|\zeta,j) + o(\Delta).$$

But for $1 \le i \le K-j$, we have

$$Q_i(t+\Delta|\zeta,j) = \big(1-\lambda_{j+i}(t)\Delta\big)Q_i(t|\zeta,j) + \lambda_{j+i-1}(t)\Delta\cdot Q_{i-1}(t|\zeta,j) + o(\Delta),$$

where $\lambda_K(t) = 0$ for all $t \in [0,T]$. Thus, the differential equations can be written as follow:

$$\frac{dQ_0(t|\zeta,j)}{dt} = -\lambda_j(t)Q_0(t|\zeta,j). \tag{4.1}$$

$$\frac{dQ_i(t|\zeta,j)}{dt} = -\lambda_{j+i}(t)Q_i(t|\zeta,j) + \lambda_{j+i-1}(t)Q_{i-1}(t|\zeta,j), \text{ for } 1 \le i \le K-j. \tag{4.2}$$

The boundary conditions are as follow:

$$\sum_{i=0}^{K-j} Q_i\,(t|\zeta,j) = 1, \text{for all } t \ge \zeta; \tag{4.3}$$

$$Q_0(\zeta|\zeta,j) = 1. \tag{4.4}$$

The unique solution to the set of Equations (4.1)-(4.4) is given by the next proposition. In this paper, all mathematical proofs are in Appendix B.1.

Proposition 1. $Q_i(t|\zeta,j)$, *with* $0 \le i \le K-j$ *and* $t \in [\zeta,T]$, *can be calculated recursively as follow:*

$$Q_0(t|\zeta,j) = e^{-\int_\zeta^t \lambda_j(s)ds} \tag{4.5}$$

$$Q_i(t|\zeta,j) = e^{-\int_\zeta^t \lambda_{j+i}(s)ds}\left[\int_\zeta^t e^{\int_\zeta^s \lambda_{j+i}(x)dx}\,\lambda_{j+i-1}(s)Q_{i-1}(s|\zeta,j)ds\right], i = 1,\cdots,K-j. \tag{4.6}$$

Furthermore, let $p(\cdot,\mu)$ be the probability mass of the Poisson distribution with rate μ, and define $P(x,\mu) = \sum_{r=x}^{\infty} p\,(r,\mu)$ (for $x = 0,1,2,\cdots$). Define $\Theta^{BD}(a,b) = \int_a^b \theta^{BD}\,(t)dt$, where $0 \le a \le b \le T$. The expected total sales under the fixed threshold policy (with j being the threshold) will be

$$S^*(j) = \sum_{i=0}^{K} i \cdot Q_i(T|0, j=0), \text{ for } j = 0;$$

$$S^*(j) = \sum_{i=0}^{K-j} \int_0^T (j+i) Q_i(T|\zeta, j) \theta^{BD}(\zeta) p\big(j-1, \Theta^{BD}(0,\zeta)\big) d\zeta + \sum_{x=0}^{j-1} x \cdot p\big(x, \Theta^{BD}(0,T)\big), \quad j = 1,2,\cdots,K \tag{4.7}$$

The explanation for Equation (4.7) is as follows. In the first term, $\theta^{BD}(\zeta)p(j-1, \Theta^{BD}(0,\zeta))$ is the probability density that the j-th sale will occur at $[\zeta - d\zeta, \zeta]$ given that the average demand rate is $\theta^{BD}(t)$ before the disclosure of inventory information, and $Q_i(T|\zeta, j)$ is the probability mass that there will be i additional sales in the remaining sale period. Hence, the first term is the expected total sales during the entire sale period, given that the number of sales will be greater than or equal to the threshold (i.e., j). The second term gives the expected total sales, given that the number of sales never reaches the threshold before the end of the sale period.

Define $\Gamma(T|\zeta, j)$ as the expected sales during $[\zeta, T]$ in addition to the j sales that have been made within $[0, \zeta]$. That is, $\Gamma(T|\zeta, j) = \sum_{i=0}^{K-j} i \cdot Q_i(T|\zeta, j)$. Then, $S^*(j)$ can be re-written as

$$S^*(j) = \Gamma(T|0, j=0), \quad \text{for } j = 0;$$

$$S^*(j) = \int_0^T \Gamma\,(T|\zeta, j) \theta^{BD}(\zeta) p\big(j-1, \Theta^{BD}(0,\zeta)\big) d\zeta + \sum_{x=0}^{j-1} x \cdot p\big(x, \Theta^{BD}(0,T)\big) + j \cdot P\big(j, \Theta^{BD}(0,T)\big), \quad j = 1, \cdots, K$$

As mentioned before, two special cases of the fixed threshold policy are worth noting, which correspond to two commonly used simple practices in the real world.

- *Disclose inventory from the very beginning of the sale*, namely, the "*always disclose*" policy. This is indeed the case when $j = 0$. Thus, the expected number of total sales is $\Gamma(T|0, j=0) = \sum_{i=0}^{K} i \cdot Q_i(T|0, j=0)$, where $Q_i(T|0, j=0)$ can be calculated according to Proposition 1.

- *Do not disclose inventory during the entire sale period*, namely, the "*never disclose*" policy. This policy is similar to a special case of setting $j = K$ in the fixed threshold policy, with the only exception that the average demand rate without the inventory information is $\theta^{ND}(t)$ instead of $\theta^{BD}(t)$. As noted earlier, the subtle difference between these two demand rates lies in whether the inventory information may be available in the future (not for the "ND" scenario but yes for the "BD" scenario). Note that the expected total sales under the "never disclose" policy reduces to $\sum_{x=0}^{K-1} x \cdot p(x, \Theta^{ND}(0,T)) + K \cdot P(K, \Theta^{ND}(0,T))$.

In general, a closed form of $Q_i(t|\zeta, j)$ is not available. However, the next proposition shows that $Q_i(t|\zeta, j)$ has a closed form when $\lambda_i(t)$ is multiplicatively separable in i and t.

Proposition 2. *Suppose the inventory-dependent demand rate follows a specific form,* $\lambda_i(t) = a(i)\lambda_0(t)$, *where* $a(0) = 1$ *and* $i \in \{0,1,\cdots,K\}$. *Without loss of generality, let* $a(K) = 0$ *such that* $\lambda_K(t) = 0$ *for any* $t \in [0,T]$. *Define* $\Lambda(\zeta, t) = \int_\zeta^t \lambda_0(s)ds$. *Then,*

$$Q_0(t|\zeta, j) = e^{-a(j)\Lambda(\zeta,t)};$$

$$Q_i(t|\zeta, j) = \left(\prod_{k=0}^{i-1} a(j+k)\right)\left(\sum_{k=0}^{i} \frac{e^{-a(j+k)\Lambda(\zeta,t)}}{\prod_{0\le l\le i, l\ne k}(a(j+l) - a(j+k))}\right), \tag{4.8}$$

$$i \in \{1, \cdots, K-j\}$$

Furthermore, given that there have been j *sales until the current time* ζ *and inventory is disclosed, the expected number of additional sales during* $[\zeta, T]$, *i.e.,* $\Gamma(T|\zeta, j) = \sum_{i=0}^{K-j} i \cdot Q_i(T|\zeta, j)$, *will be*

$$\Gamma(T|\zeta, j) = \sum_{i=0}^{K-j}\left\{e^{-a(j+i)\Lambda(\zeta,t)}\left[\sum_{k=i}^{K-j} k\left(\prod_{m=0}^{k-1} a(j+m)\right)\frac{1}{\prod_{0\le l\le k, l\ne i}(a(j+l) - a(j+i))}\right]\right\}. \tag{4.9}$$

.

Based on results of Proposition 1 (or Proposition 2 when $\lambda_i(t)$ follows the above specific form), the expected total sales for any given threshold $j \in \{0,1,\cdots,K\}$ can be calculated efficiently. Furthermore, we can employ a univariate search strategy to find out the optimal threshold, j^*.

Specifically, let $\Delta f(j) = f(j) - f(j-1)$, where $f(\cdot)$ is any function of j. Note that, for any $j \geq 1$,

$$\Delta\left(\sum_{x=1}^{j-1} x \cdot p(x, \Theta^{BD}(0,T))\right) = (j-1)p\big(j-1, \Theta^{BD}(0,T)\big);$$

$$\Delta(jP(j, \Theta^{BD}(0,T))) = P(j, \Theta^{BD}(0,T)) - (j-1)p(j-1, \Theta^{BD}(0,T)).$$

Therefore, we obtain

$$\Delta S^*(j) = \int_0^T \Delta\left(\Gamma(T|\zeta, j)p(j-1, \Theta^{BD}(0,\zeta))\right)\theta^{BD}(\zeta)d\zeta + P(j, \Theta^{BD}(0,T)), j = 1,2,\cdots,K. \tag{4.10}$$

For any $j \in \{1, \cdots, K-1\}$, it is a local maximum if $\Delta S^*(j) \geq 0$ and $\Delta S^*(j+1) \leq 0$. Finally, by comparing all the local maxima to the two boundary cases (i.e., $j = 0$ and $j = K$), the optimal threshold j^* can be easily determined. In summary, we have devised an efficient algorithm to determine the optimal inventory-level threshold for the fixed threshold policy.

4.1.2 *Time-Dependent Threshold Policy*

In this subsection, we analyze the time-dependent threshold policy. The idea is to make the disclosure decision based on not only the amount of inventory sold but also the time elapsed.

- **The time-dependent threshold policy**: *Suppose there have been $j \in \{0,1,\cdots,K\}$ sales until the current time $t \in [0,T]$, the platform should disclose its inventory if and only if t is smaller than a time threshold t_j^*, which is a decision variable to be optimized.*

In practice, the proposed time-dependent threshold policy can be implemented as follows. As time progresses, the platform monitors the dropping inventory levels. For any current inventory level (e.g., $K - j$), a time threshold will be calculated (i.e., t_j^*) and compared to the current time (i.e., t). The platform should start to broadcast its inventory if the current time is earlier than that time threshold. The intuition behind the proposed policy is that, *if inventory levels drop faster than expected*, the platform should disclose its inventory, signaling to customers that the product is of high value and will sell out soon. Thus, the disclosure of the inventory information will likely boost the future sales due to the observational learning and scarcity effects. Indeed, we can show that the proposed policy is optimal provided that the following two assumptions hold.

Assumption 1. *Once the inventory information is disclosed, it will remain published.*

This assumption is based on an accepted norm in the industry followed by all platforms who offer flash or lightning deals. To our best knowledge, we cannot find any platform flip-flopping on its information disclosure decisions. One can also argue that flip-flopping will generate customer ill-will toward the platform as it creates an impression that the platform is gaming the market. For example, eBay's inventory disclosure policy resembles a fixed-threshold policy where available inventory (or amount sold) is disclosed once the inventory level drops below a pre-set threshold, and then the inventory information remains published, thus satisfying this assumption.

Assumption 2. *The demand adjustment factor is increasing in the amount sold and decreasing in the time elapsed; that is, the demand adjustment factor increases with the past sales rate.*

The second assumption is made to be in line with the empirical findings of Cui et al. (2019), which claimed "a decrease in product availability causally attracts more sales in the future; in particular, a 10% increase in past claims leads to a 2.08% increase in cart add-ins in the next hour." Mathematically, this assumption indicates that $\beta(i, t)$ is increasing in i and decreasing in t. That is, if the inventory depletes faster, then broadcasting the remaining inventory can send a positive signal to the customers that the product is a "hot item," thus encouraging the customers to buy. Note that, in Section 4.2 we will present a detailed customer demand model under the Bayesian updating framework, which gives a specific form of the demand rates that satisfies this assumption.

Theorem 1. *Under Assumptions 1 and 2, the proposed time-dependent threshold policy is the optimal policy. That is, there exist t_j^*'s ($j = 0,1,\cdots,K$) such that the platform should disclose its inventory whenever the j-th sale occurs before t_j^*. Moreover, t_j^*'s are non-decreasing in j.*

Next, we obtain the optimal time-dependent thresholds, t_j^*'s. Suppose that there have been j sales made until the current time t, when the platform has not yet disclosed its inventory. Let $S^*(T|t,j)$ denote the optimal expected "sales-to-go" during $[t,T]$. Let $V^*(T|t,j)$ denote the expected "sales-to-go" during $[t,T]$, given that the platform decides not to disclose its inventory now (but it may disclose in the future). In contrast, recall that $\Gamma(T|t,j)$ is defined as the expected "sales-to-go" during $[t,T]$, given that the platform broadcasts its inventory immediately from t and henceforward. Therefore, the optimization problem can be expressed as follows:

$$S^*(T|t,j) = \max\{\Gamma(T|t,j), V^*(T|t,j)\}. \tag{4.11}$$

We start the analysis of Equation (4.11) by showing the first two steps, $j = K-1$ and $j = K-2$, and then generalize these steps for any other $j \in \{0,1,\cdots,K-3\}$. First, when $j = K$, we can set $t_K^* = T$ without loss of generality, so we start from $j = K-1$. The goal is to find $t_{K-1}^* \in [0, t_K^*] = [0,T]$.

Let $q_1(y|t,K-1)$ denote the probability density function of the following event: given that there have been $K-1$ sales made until t, the disclosure of inventory will be at $[y-dy, y]$ when the first future demand occurs. Then, $V^*(T|t,K-1)$ can be written as

$$V^*(T|t,K-1) = \int_t^{t_K^*=T} [\,1 + \Gamma(T|y,K)]q_1(y|t,K-1)dy. \tag{4.12}$$

Note that $q_1(y|t,K-1) = \theta^{BD}(y)p(0,\Theta^{BD}(t,y))$ and $\Gamma(T|y,K) = 0$. Thus, Equation (4.12) becomes

$$V^*(T|t,K-1) = \int_t^{t_K^*=T} \theta^{BD}(y)p(0,\Theta^{BD}(t,y))dy = P(1,\Theta^{BD}(t,T)).$$

This makes sense as, with probability $P(1, \Theta^{BD}(t,T)) = \sum_{x=1}^{\infty} p\,(x, \Theta^{BD}(t,T))$, there will be one more sale during $[t,T]$ given that there is only one item left at t.

Hence, t_{K-1}^{*} will be the largest value of t such that $\Gamma(T|t, K-1) \geq V^{*}(T|t, K-1) = P(1, \Theta(t,T))$; that is, the expected "sales-to-go" if the platform discloses its inventory at t is greater than or equal to the expected "sales-to-go" if the platform does not disclose its inventory at that time.

Next, suppose we have obtained t_{K-1}^{*} and consider $j = K-2$. The goal is to find $t_{K-2}^{*} \in [0, t_{K-1}^{*}]$. Similar to the definition of $q_1(y|t, K-1)$, let $q_i(y|t, K-2)$ denote the probability density function of the following event: given that there have been $K-2$ sales until t, the disclosure of inventory will be made at $[y-dy, y]$ when the i-th future demand occurs, where $i \in \{1,2\}$. Moreover, define $r_i(y|t, K-2)$ as the probability of the following event: given that there have been $K-2$ sales until t, there will be i additional sales and the platform will not broadcast its inventory during $[t, y]$, where $i \in \{1,2\}$. Then, $V^{*}(T|t, K-2)$ can be written as

$$\begin{aligned} V^{*}(T|t, K-2) = &\int_{t}^{t_{K-1}^{*}} [1 + \Gamma(T|y, K-1)]q_1(y|t, K-2)dy \\ &+ \int_{t_{K-1}^{*}}^{T} [2 + \Gamma(T|y, K)]q_2(y|t, K-2)dy + r_1(T|t, K-2). \end{aligned} \tag{4.13}$$

In Equation (4.13), the first term is the expected future sales, providing that the following scenario happens: The first future demand (i.e., the $(K-2+1)$-th sale) will occur at $[y-dy, y]$ and the platform will start to disclose its inventory from then on, which requires $y \leq t_{K-1}^{*}$. In that case, the expected sales during $[t,T]$ will be $1 + \Gamma(T|y, K-1)$. The second term can be explained in a similar way. But, since the platform will disclose its inventory at time $[y-dy, y]$ upon the occurrence of the second future demand (i.e., the $(K-2+2)$-th sale), we must have $y > t_{K-1}^{*}$ because, otherwise, the platform would have disclosed its inventory upon the occurrence of the first future demand (i.e., the $(K-2+1)$-th sale). The third term gives the expected sales from t

to T given that the platform will not disclose its inventory until T. In that case, demand during $[t, T]$ cannot be greater than or equal to two items as, otherwise, the platform would have disclosed its inventory (note that $t_K^* = T$).

Furthermore, for any $t \leq t_{K-1}^*$,

$$q_1(y|t, K-2) = r_0(t|t, K-2)\theta^{BD}(y)p(0, \Theta^{BD}(t, y)) = \theta^{BD}(y)p(0, \Theta^{BD}(t, y));$$
$$q_2(y|t, K-2) = r_0(t_{K-1}^*|t, K-2)\theta^{BD}(y)p\big(1, \Theta^{BD}(t_{K-1}^*, y)\big).$$

Here, we provide an explanation for the expression of $q_2(y|t, K-2)$. In order to have the platform disclose its inventory upon the occurrence of the second future demand at $[y-dy, y]$, two conditions are required: (1) There will be no demand during $[t, t_{K-1}^*]$ because, otherwise, the platform would have disclosed its inventory upon the occurrence of the first future demand; the probability of such a condition is $r_0(t_{K-1}^*|t, K-2)$. (2) There will be demand for two items during $(t_{K-1}^*, y]$, and, in particular, the second future demand will occur at $[y-dy, y]$; the probability density of such a condition is $\theta^{BD}(y)p(1, \Theta^{BD}(t_{K-1}^*, y))$.

Therefore, calculation of Equation (4.13) reduces to a problem of calculating $r_0(t_{K-1}^*|t, K-2)$ and $r_1(T|t, K-2) = r_1(t_K^*|t, K-2)$, which actually have closed-form expressions. Now,

$$r_0(t_{K-1}^*|t, K-2) = p\big(0, \Theta^{BD}(t, t_{K-1}^*)\big)r_0(t_{K-1}^*|t_{K-1}^*, K-2) = p\big(0, \Theta^{BD}(t, t_{K-1}^*)\big);$$

$$r_1(t_K^*|t, K-2) = p(0, \Theta^{BD}(t, t_{K-1}^*))r_1(t_K^*|t_{K-1}^*, K-2) = p(0, \Theta^{BD}(t, t_{K-1}^*))p(1, \Theta^{BD}(t_{K-1}^*, t_K^*))$$

Finally, once the calculation of Equation (4.13) is done, t_{K-2}^* will be the largest value of t such that $\Gamma(T|t, K-2) \geq V^*(T|t, K-2)$. Now, we have completed the calculation for $j = K-2$.

The above analysis concerning $j = K-1$ and $j = K-2$ can be generalized for any other $j \in \{0,1,\cdots, K-3\}$. The following two general definitions will be used in the next proposition.

- Let $q_i(y|t,j)$ denote the probability density function of the following event: given that there are j sales and the platform has not disclosed its inventory until t, it will disclose its inventory at $[y-dy,y]$ when the i-th additional future demand occurs, where $i \in \{1,\cdots,K-j\}$.

- Let $r_i(y|t,j)$ denote the probability of the following event: given that there are j sales and the platform has not disclosed its inventory until t, there will be i additional sales and the platform will not disclose its inventory during $[t,y]$, where $i \in \{1,\cdots,K-j\}$.

It is important to note that $q_i(y|t,j)$ can be written as follows:

For $i=1, q_1(y|t,j)=\theta^{BD}(y)p\big(0,\Theta^{BD}(t,y)\big)$;

$$\text{For } 1<i\le K-j, q_i(y|t,j)=\sum_{n=0}^{i-2} r_n\left(t^*_{j+i-1}\middle|t,j\right)\theta^{BD}(y)p\left(i-1-n,\Theta^{BD}\left(t^*_{j+i-1},y\right)\right). \quad (4.14)$$

In Equation (4.14), with probability $r_n(t^*_{j+i-1}|t,j)$, there will be n additional sales and the platform will not broadcast its inventory during $[t,t^*_{j+i-1}]$. Note that n cannot be equal to or greater than $i-1$ as, otherwise, the platform would have broadcast its inventory. Next, with probability density $\theta^{BD}(y)p(i-1-n,\Theta^{BD}(t^*_{j+i-1},y))$, there will be $i-n$ more sales in $(t^*_{j+i-1},y]$, and, in particular, the $(i-n)$-th additional sale since t^*_{j+i-1} (i.e., the i-th additional sale since t) will occur at $[y-dy,y]$.

Moreover, for any given j, consider $1\le k\le K-j$ and $0\le n\le k-1$. Note that

$$r_n(t^*_{j+k}|t,j)=p(0,\Theta^{BD}(t,t^*_{j+1}))r_n(t^*_{j+k}|t^*_{j+1},j), \quad (4.15)$$

where $r_n(t^*_{j+k}|t^*_{j+1},j)$ can be calculated recursively as follow: ($1\le k\le K-j$ and $0\le n\le k-1$)

For $k=1, r_0(t^*_{j+k}|t^*_{j+1},j)=1$;

$$\text{For } k \geq 2, r_n(t_{j+k}^*|t_{j+1}^*, j) = \sum_{m=0}^{n \wedge (k-2)} r_m\,(t_{j+k-1}^*|t_{j+1}^*, j) p(n-m, \Theta^{BD}(t_{j+k-1}^*, t_{j+k}^*)). \tag{4.16}$$

In Equation (4.16), with probability $r_m(t_{j+k-1}^*|t_{j+1}^*, j)$, there will be $m \leq n$ additional sales and the platform will not disclose its inventory in $[t_{j+1}^*, t_{j+k-1}^*]$. Note that m cannot be equal to or greater than $k-1$ as, otherwise, the platform would have disclosed its inventory before t_{j+k-1}^*. Moreover, with probability $p(n-m, \Theta^{BD}(t_{j+k-1}^*, t_{j+k}^*))$, there will be $n-m$ more sales in $(t_{j+k-1}^*, t_{j+k}^*]$. Put these probabilities together, Equation (4.16) gives the probability that there will be n additional sales and the platform will not disclose its inventory in $[t_{j+1}^*, t_{j+k}^*]$. As indicated, n cannot exceed $k-1$; otherwise, the platform would have disclosed its inventory before t_{j+k}^*.

For instance, for $j = K-2$, Equation (4.14) reduces to $q_1(y|t, K-2) = \theta^{BD}(y) p(0, \Theta^{BD}(t, y))$ and $q_2(y|t, K-2) = r_0(t_{K-1}^*|t, K-2)\theta^{BD}(y) p(1, \Theta^{BD}(t_{K-1}^*, y))$. Moreover, Equation (4.15) and Equation (4.16) together lead to $r_0(t_{K-1}^*|t, K-2) = p(0, \Theta^{BD}(t, t_{K-1}^*)) r_0(t_{K-1}^*|t_{K-1}^*, K-2) = p(0, \Theta^{BD}(t, t_{K-1}^*))$ when $k = 1$ and $n = 0$, and $r_1(t_K^*|t, K-2) = p\big(0, \Theta^{BD}(t, t_{K-1}^*)\big) r_1(t_K^*|t_{K-1}^*, K-2) = p\big(0, \Theta^{BD}(t, t_{K-1}^*)\big) p(1, \Theta^{BD}(t_{K-1}^*, t_K^*))$ when $k = 2$ and $n = 1$. These results are consistent with what we have derived for the case of $j = K-2$.

Proposition 3. *Under the proposed time-dependent threshold policy, t_j^* ($j \in \{0,1,\cdots,K-1\}$) can be calculated as follow: Suppose t_k^* ($k = j+1,\cdots,K$) have all been determined, then t_j^* is the largest value of t such that $\Gamma(T|t,j) \geq V^*(T|t,j)$. Specifically,*

$$\begin{aligned} V^*(T|t,j) = \int_t^{t_{j+1}^*} [1 + \Gamma(T|y, j+1)] q_1(y|t,j) dy + \sum_{i=2}^{K-j} \int_{t_{j+i-1}^*}^{t_{j+i}^*} [i + \Gamma(T|y, j+i)] q_i(y|t,j) dy \\ + \sum_{n=0}^{K-j-1} n \cdot r_n(T|t,j), \end{aligned} \tag{4.17}$$

where $q_i(y|t,j)$ ($i = 0,1,\cdots,K-j$) and $r_n(t_{j+k}^*|t,j)$ ($1 \leq k \leq K-j$ and $0 \leq n \leq k-1$) are determined by Equations (4.14)-(4.16).

One can verify that, when $j = K - 1$, Equation (4.17) reduces to $V^*(T|t, K-1) = \int_t^{t_K^*}[1 + \Gamma(T|y, K)]q_1(y|t, K-1)dy$, which is indeed Equation (4.12). When $j = K - 2$, Equation (4.17) reduces to $\int_t^{t_{K-1}^*}[1 + \Gamma(T|y, K-1)]q_1(y|t, K-2)dy + \int_{t_{K-1}^*}^{t_K^*}[2 + \Gamma(T|y, K)]q_2(y|t, K-2)dy + r_1(T|t, K-2)$, which is Equation (4.13).

4.2 A Model of Customers' Propensity of Purchase

In the previous section, we have derived the optimal thresholds for the fixed and time-dependent threshold policies based on a general functional form of the average demand rates. In the literature on marketing and economics, the demand function can take various specific forms (e.g., the Cobb-Douglas utility function), so the question is: what specific functional form of the demand rates can we employ in the numerical validation of our models (to be presented in the next section)?

In the context of lightning or flash deals, there are herding and scarcity effects on the customers' purchasing behavior, as evidenced by Cui et al. (2019): "by observing past purchasing decisions from inventory information, customers can draw inferences and update their beliefs about deal quality, especially when their prior knowledge about the deal is imperfect. A product's low inventory level can also create an out-of-stock pressure among customers, prompting a resulting urgency to buy that product immediately." In this vein, we develop a detailed customer demand model under the Bayesian updating framework to capture the two salient effects in online lightning or flash deals, which will be employed in our numerical study in the next section.

First, customers' propensity of purchase increases with the *herding effect*. This phenomenon is well supported by empirical evidence. Zhang and Liu (2012) explained this phenomenon through the mechanism of rational herding or observational learning. We follow the extant literature to model this observational learning process using the Bayesian belief updating framework (Erdem and Keane 1996). In the context of our study, the true product quality can be inferred from the information provided on the website, such as online reviews and customer ratings which are provided, for example, for Amazon flash sales. Previous studies (e.g., Ho et al. 2017) showed that online reviews, with a sufficient volume such as the ones on Amazon, provide accurate information about the true product quality. However, there are still uncertainties which prevent customers from immediately trusting the true product quality, and instead they go through a process to update their perceived quality. That is, a potential buyer starts with her own prior about the quality (or value) of a product but receives signals of the true quality through observing others' purchasing behavior. A purchase incidence by other customers is considered as a confirmation of the true quality and sends a signal carrying this confirmation, thus increasing the perceived quality, and consequently the purchase propensity, of the undecided customers.

Second, customers' propensity of purchase is stimulated by the *scarcity effect*. That is, customers have a higher purchase propensity when the product is more scarce. Previous studies have documented increased desirability among consumers to purchase in various markets adopting a marketing strategy of scarcity (Eisend 2008, Jung and Kellaris 2004, Yang et al. 2020). The quantity scarcity of a product could be a consequence of either excess demand or limited supply, depending on individuals' underlying motivational goal (Ku et al. 2012). In our setting, customers can be prevention-motivated buyers who prefer demand-induced scarce products; they purchase a product that has already attracted a considerable number of buyers, which allows them to have a sense of security, as they believe that it is less risky to follow others' decisions or simply feel more assured to be part of a majority group (Corazzini and Greiner 2007).

In summary, the premise of our demand model is well supported by the extant literature that a customer's propensity of purchase is increasing in not only the perceived quality but also the scarcity of the product. Specifically, let q be the *perceived quality* and p_{so} be the *perceived probability* that the product will be sold out. The herding behavior of the customer updates q according to the Bayes's rule, and p_{so} captures the scarcity of the product. We assume that customers have heterogeneous prior beliefs about the product's true quality and are segmented by their prior beliefs. Therefore, for a customer whose perceived quality and scarcity of the product are q and p_{so}, respectively, her propensity of purchasing the product at cost c, can be defined as follows:

$$U = \begin{cases} (b_1 q + b_2 p_{so} - c) + \varepsilon, & \text{if buy} \\ 0, & \text{otherwise} \end{cases}$$

where b_1 and b_2 are coefficients associated with the perceived quality and scarcity of the product, and the random component ε captures the idiosyncratic preference of the customer for the product, which we assume follows a logistic distribution with mean zero and scale ω. This is a binary-logit model, where the cumulative distribution function of ε is

$$\text{Prob}\{\varepsilon \leq x\} = \frac{1}{1 + e^{-x/\omega}}.$$

Then, the probability of the customer buying the product is

$$\text{Prob}\{U \geq 0\} = 1 - \frac{1}{1 + e^{(b_1 q + b_2 p_{so} - c)/\omega}} = \frac{1}{e^{-(b_1 q + b_2 p_{so} - c)/\omega} + 1} \tag{4.18}$$

To capture the updating of the customer's perceived quality, we adopt a Bayesian updating framework which is in line with DeGroot (2005, Chapter 9.5). Suppose that a customer from segment x has a prior belief of the product's uncertain quality, $\tilde{q}_{x,0}$, which follows a normal distribution with mean $q_{x,0}$ and variance σ_0^2. Without loss of generality, we assume that $q_{x,0} \in [q_{min}, q_{max}]$ and follows a distribution $H(y) = \text{Prob}\{q_{x,0} \leq y\}$.

Next, we will distinguish the customer's information in three different scenarios:

- The customer knows that the platform will never disclose inventory information.

- The customer knows that the platform employs a threshold policy, while the platform has not disclosed its inventory (i.e., the threshold for disclosure has yet reached).
- The platform publishes its inventory information (i.e., the threshold has been reached).

In the first scenario, where inventory information is never disclosed, the customer cannot update the prior belief through an observation of the inventory levels. In such a case, the perceived quality of the product is equal to the mean of the customer's prior distribution, $q_{x,0}$. Also, the customer holds a certain belief about the probability that the product will be sold out, denoted by p_{so}^{ND}, whereby "*ND*" stands for "never disclose," and the customer cannot update this belief as the platform will never provide information about the inventory availability. Thus, the purchase propensity of the customer under the first scenario can be written as

$$U_x^{ND} = \left(b_1 q_{x,0} + b_2 p_{so}^{ND} - c\right) + \varepsilon. \tag{4.19}$$

In the second scenario, where the platform employs a threshold policy but the threshold for disclosure has not been reached, the perceived quality of a customer from segment x is $q_{x,0}$, as the customer does not possess information about the inventory sold to update her belief. However, in contrast to the first scenario, now the customer knows that the inventory information is not available simply because the threshold for disclosure has not been reached, which indeed signals that there is still plenty of inventory left. Thus, the customer's belief of the sell-out probability, p_{so}^{BD}, where "*BD*" stands for "before-disclosure," can be different from p^{ND}. As noted earlier, one may argue that $p_{so}^{BD} \leq p_{so}^{ND}$ as the customer will interpret no disclosure under a threshold policy as a signal for slow sales. The purchase propensity of the customer under the second scenario is

$$U_x^{BD} = \left(b_1 q_{x,0} + b_2 p_{so}^{BD} - c\right) + \varepsilon. \tag{4.20}$$

In the third scenario, the customer's perceived quality and scarcity of the product can be updated upon observation of the inventory information (i.e., past sales and available inventory).

As mentioned above, previous studies (Erdem and Keane 1996, Zhang and Liu 2012, Ho et al. 2017) implied that online reviews, with a sufficient volume such as the ones on Amazon, provide accurate signals about the true product quality; however, there are uncertainties which prevent customers from immediately trusting the true product quality, and instead they remain skeptical about the true product quality and go through a process to update their perceived quality. Let the uncertain true quality of the product, $\tilde{q}_T$, follow a normal distribution with mean q_T and variance σ_T^2; for example, q_T is the average customer ratings of the product on the website and σ_T^2 is a parameter pertaining to the precision of that signal. Importantly, observational learning (herding) happens when a purchase incidence by other customers is considered as a confirmation of the quality signal, and the purchase incidence sends a signal carrying this confirmation. As a result, the increasing incidence of purchases (i.e., more past sales) stimulates a potential buyer's propensity of purchase.

To reflect the herding behavior of customers, we assume that the customer starts with her own prior about the quality of the product, $q_{x,0} \leq q_T$, such that the customer's perception of the product quality increases with the number of units sold. This assumption also implies that initially, the customer does not immediately trust the quality signal without reserve, but, as the number of units sold increases, the customer updates her belief and views the quality signal more credible. Thus, with the observation of the inventory sold, the customer updates her belief based on the Bayesian updating rule (DeGroot 2005, Chapter 9.5). After observing i sales, the customer's posterior belief $\tilde{q}_x(i)$ is updated to a normal distribution with mean $q_x(i)$ and variance σ_i^2, where

$$q_x(i) = \frac{\frac{1}{\sigma_0^2} q_{x,0} + \frac{i}{\sigma_T^2} q_T}{\frac{1}{\sigma_0^2} + \frac{i}{\sigma_T^2}} \text{ and } \sigma_i^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{i}{\sigma_T^2}},$$

where $1/\sigma_0^2$ and $1/\sigma_T^2$ are referred to as the precision of the customer's prior belief and the signal of the product's quality, respectively (DeGroot 2005, Chapter 9.5). Furthermore, recall that i sales translates to the fraction of inventory sold, $s(i) = i/K$, and time t translates to the fraction of the sale period elapsed, $\tau(t) = t/T$. By a slight abuse of notation, let s and τ represents the faction of inventory sold and the fraction of sale period elapsed, respectively. Then, we have

$$q_x(i) \triangleq q_x(Ks) = \frac{\frac{1}{\sigma_0^2 K}}{\frac{1}{\sigma_0^2 K} + \frac{s}{\sigma_T^2}} q_{x,0} + \frac{\frac{s}{\sigma_T^2}}{\frac{1}{\sigma_0^2 K} + \frac{s}{\sigma_T^2}} q_T = (1 - \rho(s))q_{x,0} + \rho(s)q_T. \tag{4.21}$$

As can be seen, the customer's perceived quality of the product after updating, $q_x(i)$, can be viewed as a weighted sum of the mean of the customer's prior belief of the product quality, $q_{x,0}$, and the mean signal of the product quality, q_T (e.g., the average customer ratings on the website). Moreover, $\rho(s)$ is an increasing function of s. That is, upon observing more sales and less inventory left, the customer views the signal of the true product quality (e.g., average customer ratings) more credible, thus increasing the customer's propensity to buy, which reflects the herding effect.

Next, we turn to the customer's estimation of the product's scarcity. Upon observation of i sales at time t, the customer's estimation of the sell-out probability is updated to $p_{so}(s, \tau)$, which depends on the fraction of inventory sold, $s = i/K$, and the fraction of time elapsed, $\tau = t/T$. In general, if sales go faster (s increases or τ decreases), the sell-out probability must increase. We recognize that there are alternative ways of updating the sell-out probability. Below are two examples.

- Note that the customer observes the average sales rate, $i/t = (Ks)/(T\tau)$. If the customer views the sales rate as constant and demand as deterministic, she will expect the product to be sold out if demand in the remaining time of the sales, $(i/t)(T-t)$, exceeds the inventory left, $K-i$. In other words, $p_{so}(s,\tau)$ can take the following form:

$$p_{so}(s,\tau) = \begin{cases} 1, & \text{if } \dfrac{i}{t}(T-t) \geq K-i;\ \text{that is, } \dfrac{s}{1-s} \geq \dfrac{\tau}{1-\tau} \\ 0, & \text{otherwise} \end{cases} \tag{4.22}$$

- If the customer is sophisticated enough to deal with random demand, she can infer the probability that demand in the remaining time of the sales exceeds the inventory left, while only assuming that the sales rate is constant. For instance, the customer may assume that demand in the remaining sale period, $D([t,T])$, follows a distribution which has a cumulative distribution function $F(x|[t,T])$ with mean $(i/t)(T-t)$. Then, $p_{so}(s,\tau)$ is given by

  $$p_{so}(s,\tau) = \text{Prob}\{D([t,T]) \geq K-i\} = 1 - F((K-i)|[t,T]) = 1 - F(K(1-s)|T \times [\tau, 1]),$$

  With that said, we would like to point out that most customers in the real world are not sophisticated enough, and they even do not know what distribution of demand can be. Hence, in our numerical study in Section 4.3, we will apply Equation (4.22), as it resembles the customers' intuitive estimation of the sell-out probability when they are not sophisticated enough.

In short, in the third scenario where inventory information is published, the purchase propensity of a customer from segment x can be written as

$$U_x(s,\tau) = (b_1 q_x(Ks) + b_2 p_{so}(s,\tau) - c) + \varepsilon, \tag{4.23}$$

where $q_x(Ks)$ and $p_{so}(s,\tau)$ are given by Equation (4.21) and Equation (4.22), respectively. Both $q_x(Ks)$ and $p_{so}(s,\tau)$ increase with s, and $p_{so}(s,\tau)$ decreases with τ, so $U_x(s,\tau)$ is increasing in s and decreasing in τ.

Finally, let $\Lambda(t)$ denote the average arrival rate of the customers and note that a customer will purchase the product if and only if her purchase propensity is non-negative. Then, the function of the average demand rates can be derived from aggregating the customers' purchase decisions.

1. If the customer knows that the platform will never reveal inventory information, then the average demand rate, $\theta^{ND}(t)$, is

$$\theta^{ND}(t) = \Lambda(t)\int_{q_{min}}^{q_{max}} \text{Prob}\{U_x^{ND} \geq 0\} dH\big(q_{x,0}\big)$$

$$= \Lambda(t)\int_{q_{min}}^{q_{max}} \frac{1}{e^{-(b_1 q_{x,0}+b_2 p_{so}^{ND}-c)/\omega}+1} dH(q_{x,0})$$

2. If the customer knows that the platform employs a threshold policy while so far the threshold has not been reached, then the average demand rate, $\theta^{BD}(t)$, is

$$\theta^{BD}(t) = \Lambda(t)\int_{q_{min}}^{q_{max}} \text{Prob}\{U_x^{BD} \geq 0\} dH\big(q_{x,0}\big)$$

$$= \Lambda(t)\int_{q_{min}}^{q_{max}} \frac{1}{e^{-(b_1 q_{x,0}+b_2 p_{so}^{BD}-c)/\omega}+1} dH(q_{x,0})$$

3. If the platform has disclosed the inventory information, then the average demand rate, $\lambda(i,t)$ or, equivalently, $\lambda(Ks,T\tau)$, is

$$\lambda(Ks,T\tau) = \Lambda(t)\int_{q_{min}}^{q_{max}} \text{Prob}\{U_x(s,\tau) \geq 0\} dH\big(q_{x,0}\big)$$

$$= \Lambda(t)\int_{q_{min}}^{q_{max}} \frac{1}{e^{-(b_1 q_x(Ks)+b_2 p_{so}(s,\tau)-c)/\omega}+1} dH(q_{x,0})$$

It is worth noting that the ratio of $\lambda(Ks,T\tau)$ to $\theta^{BD}(t)$, which gives the demand adjustment factor, is increasing in the fraction of inventory sold (i.e., s) and decreasing in the fraction of time elapsed (i.e., τ), thus satisfying Assumption 2.

4.3 NUMERICAL STUDY

In this section, we employ the optimization algorithms developed in Section 4.1 through a comprehensive numerical study to explore performances of the optimal fixed and time-dependent threshold policies, as compared to the two simple yet common inventory disclosure practices, "always disclose" and "never disclose." In particular, we examine impacts of the two salient effects – the observational learning (herding) and scarcity effects – on performances of these inventory disclosure policies by employing the customer's purchase propensity model developed in Section 4.2.

For the customer's purchase propensity functions (see Equations (4.19), (4.20), and (4.23)), we normalize c to 1 such that other parameters can be discussed with respect to c. In particular, the standard deviation of the random component (i.e., ε), which we denote as σ_ε, represents how much random variation that the customer's propensity of purchase can have relative to c. As ε follows a logistic distribution with scale parameter ω, we have $\sigma_\varepsilon = (\sqrt{3}/3)\pi\omega$ as the one-to-one mapping between σ_ε and ω.

In the Bayesian updating framework, recall that a customer from segment x has a prior belief of the product's quality, which follows a normal distribution with mean $q_{x,0} \in [q_{min}, q_{max}]$ and variance σ_0^2. The customer also receives a quality signal (e.g., customer reviews and ratings on the website), which follows a normal distribution with mean q_T and variance σ_T^2. Here, we normalize $q_T = 1$ and assume that $q_{x,0}$ is uniformly distributed between $q_{min} = 0$ and $q_{max} = 1$. Indeed, we do not need to specify values for σ_0^2 and σ_T^2; instead, we can use the maximum value of $\rho(s)$, i.e., $\rho_{max} = \rho(1)$, to identify the model (see Equation (4.21)). Moreover, we vary the weights associated with the observational learning and scarcity effects, b_1 and b_2, respectively. As b_1 (or

b_2) increases, the observational learning effect (the scarcity effect, respectively) becomes relatively more prominent.

Last, we assume that the average arrival rate of customers, $\Lambda(t)$, takes an exponential form with respect to $t \in [0, T]$; that is, $\Lambda(t) = \Lambda_0 e^{a_0 t}$. When $a_0 > 0$ (< 0), the average arrival rate increases (decreases) towards the end of the sale period, while $a_0 = 0$ implies that the average arrival rate is a constant Λ_0. To put the parameter Λ_0 into perspective, not only do we consider various initial inventory levels, K, we also examine various ratios of Λ_0 to K. Table 4.2 summarizes the parameter setting of the core testbed in our numerical study, which leads to 2,250 cases in total.

Table 4.2. Setting of parameters of the core testbed.

Parameter	Values of the parameter
T	1
K	{10, 20, 30}
Λ_0	$\{0.5,1.0,1.5\} \times K$
a_0	{0, 0.3}
c	1
b_1	{0.5,0.8,1.0,1.2,2.0}
b_2	{0.5,0.8,1.0,1.2,2.0}
σ_ε	0.25
ρ_{max}	0.8
$q_{x,0}$	Uniformly distributed in [0,1]
p_{so}^{ND}	{0.1,0.2,0.3,0.4,0.5}
p_{so}^{BD}	$1 \times p_{so}^{ND}$

To explore the relative performances among various policies examined in this paper, we exa mine the following relative measures. Let $S^*[X]$ denote the expected total sales under a particular policy $X \in$ {"never disclose," "always disclose," fixed threshold, time-dependent threshold}.

$$R_1 = \frac{S^*[\text{always disclose}]}{S^*[\text{never disclose}]} - 1, \qquad R_2 = \frac{S^*[\text{fixed threshold}]}{\max\{S^*[\text{always disclose}], S^*[\text{never disclose}]\}} - 1,$$

$$R_3 = \frac{S^*[\text{time-dependent threshold}]}{S^*[\text{fixed threshold}]} - 1, \quad R_4 = \frac{S^*[\text{time-dependent threshold}]}{\max\{S^*[\text{always disclose}], S^*[\text{never disclose}]\}} - 1.$$

Note that R_1 measures the improvement of the "always disclose" policy compared to the "never disclose" policy, which can be negative as the former is not necessarily better than the latter. R_2 measures the improvement of the optimal fixed threshold policy compared to the better of the two simple policies, while R_3 measures the *additional* improvement of switching from the optimal fixed threshold policy to the optimal time-dependent threshold policy. R_4 is the improvement of the optimal time-dependent threshold policy compared to the better of the two simple policies.

Table 4.3 below shows the magnitudes of the relative improvements in cases where R_2, R_3, or R_4 is strictly positive; in other words, those are cases where the fixed or time-dependent threshold policy outperforms the two simple policies. Among the total 2,250 cases, the fixed threshold policy outperforms the better of the two simple policies ($R_2 > 0$) in 1,245 cases and does equally well in the rest cases. The time-dependent threshold policy outperforms the fixed threshold policy ($R_3 > 0$) in 1,687 cases and does no worse in the rest cases. The time-dependent threshold policy outperforms the simple policies ($R_4 > 0$) in 1,973 cases and does equally well in the rest cases.

Table 4.3. Summary statistics of the relative improvements measured by R_2, R_3, and R_4

	Mean	Standard Deviation	25% quantile	50% quantile	75% quantile	Maximum
Cases with $R_2 > 0$	4.58%	5.54%	0.37%	2.18%	7.10%	29.51%
Cases with $R_3 > 0$	3.60%	4.93%	0.23%	1.59%	4.80%	28.76%
Cases with $R_4 > 0$	6.06%	7.04%	0.70%	3.32%	9.19%	33.42%

Table 4.3 shows that the optimal fixed threshold policy can achieve a significant improvement compared to the two simple policies. Since the fixed threshold policy is also easy to implement in practice, the significant improvement that it achieves is encouraging. Furthermore, even though the fixed threshold policy performs well, the optimal time-dependent threshold policy can still achieve a notable improvement as compared to the optimal fixed threshold policy. Another

intriguing observation from our numerical experimentation is that the underlying conditions for R_2, R_3, or R_4 reaching a considerable magnitude are different, which motivates us to delve deeper into the demand characteristics under which a policy significantly outperforms another.

In subsections 5.1 and 5.2, we will examine impacts of the parameters pertaining to the observational learning and scarcity effects, respectively. Our **default parameter setting** is: $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, and $T = 1$. We fix other parameters at their default values when we examine the impact of a particular parameter.

4.3.1 *Impact of the Observational Learning (Herding) Effect*

The observational learning effect plays a role after the platform discloses its inventory information, so it is relevant under the fixed and time-dependent threshold policies. In the customer's purchase propensity model developed in Section 4.2, parameters pertaining to the observational learning effect include b_1, ρ_{max} (i.e., the maximal $\rho(s)$), and the distribution of the customers' prior belief of the product quality. We now examine how these parameters affect the performance of different policies.

Impact of the weight associated with the observational learning effect in customers' purchasing decisions. Figure 4.1 shows the relative performances of various policies as b_1 varies, while other parameters being fixed at their default values. Note that a higher value of b_1 implies that the observational learning effect plays a more prominent role in customers' purchasing decisions.

First, R_1 increases with b_1. When b_1 is small, the observational learning (herding) effect has little impact on the customers' purchase decisions, which explains why "always disclose" is inferior to “never disclose” in such cases. As b_1 increases, however, customers’ purchasing decisions are more influenced by herding, which benefits the “always disclose” policy.

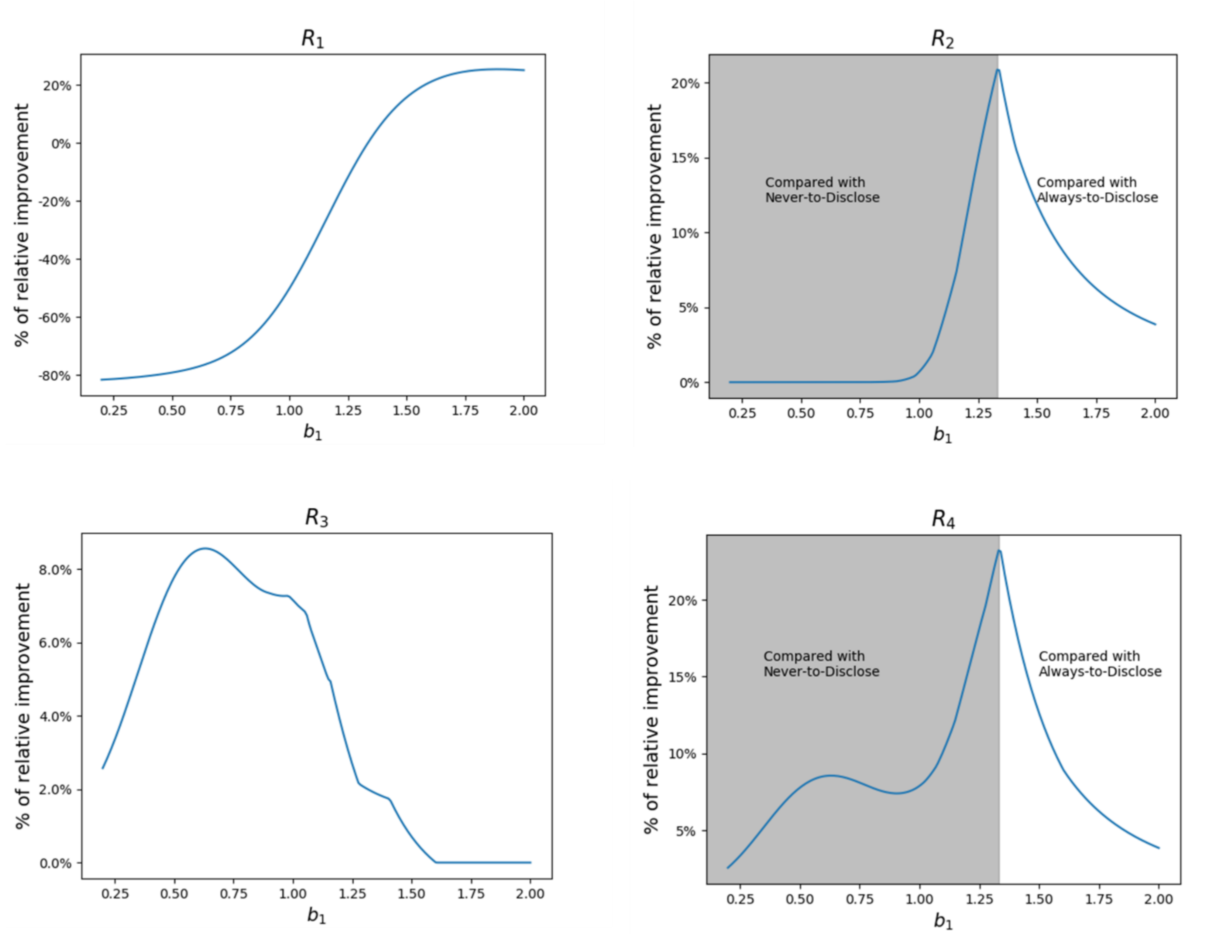

Figure 4.1. Measures of improvements as the weight attached to the observational learning effect, b_1, varies.

Note. $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

Second, we consider the optimal fixed threshold policy versus the better of the two simple policies. When b_1 is small, as the observational learning effect is weak, the advantage of the optimal threshold policy over the “never disclose” policy is insignificant, so R_2 is almost zero. As b_1 increases, the observation learning effect raises the expected sales under the optimal fixed threshold policy, so R_2 increases. Note that the peak of R_2 appears at a threshold of b_1 such that “always disclose” and “never disclose” perform equally well. As b_1 increases further, R_2 measures the improvement of the optimal fixed threshold policy as compared to the “always disclose” policy. In such cases, the observational learning effect is more prominent and benefits the optimal fixed threshold policy, while it benefits the “always disclose” policy even more, so R_2 decreases.

Third, we examine the additional improvement of switching from the optimal fixed threshold policy to the optimal time-dependent threshold policy. As can be seen, R_3 has a unimodal behavior with respect to b_1. To better analyze the behavior of R_3, Figure 4.2 shows the optimal thresholds under both the fixed and time-dependent threshold policies as b_1 varies. The vertical axis of the left graph represents the optimal fixed threshold, while the vertical axis of the right graph reveals the optimal time thresholds for different levels of inventory sold. Our observations are as below.

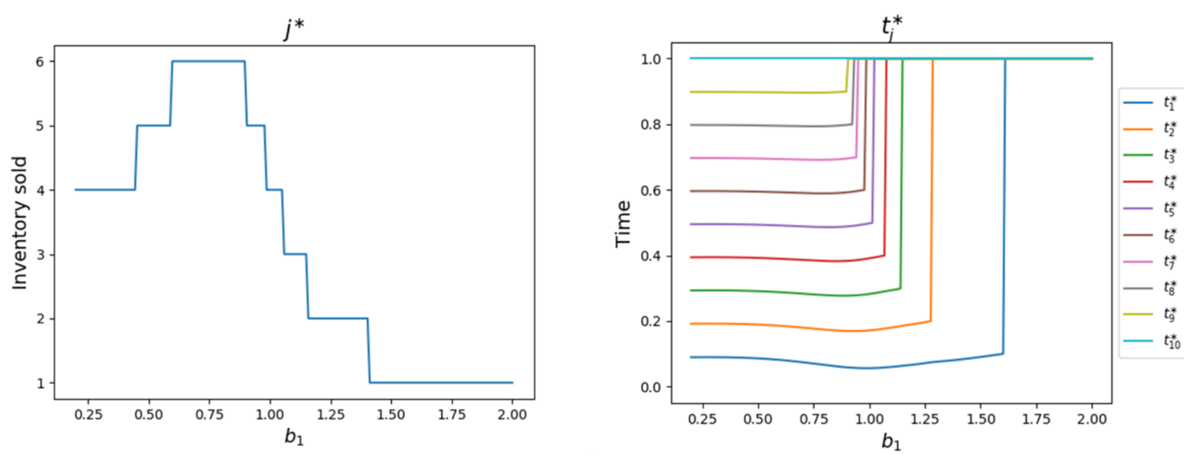

Figure 4.2. The optimal thresholds of the fixed (left) and time-dependent (right) threshold policies as b_1 varies.

Note. $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

- The optimal fixed threshold first increases then decreases as b_1 increases. When the observational learning effect is insignificant (i.e., b_1 is small), it is more desirable for the platform to withhold inventory information until a smaller fraction of inventory remains to increase the scarcity. However, as b_1 becomes moderate or large, the observational learning effect becomes the dominating factor, and the platform is motivated to disclose inventory information earlier (i.e., lower the threshold for disclosure) to benefit from the customers' herding behaviors.

- The optimal time thresholds associated with different levels of inventory sold, t_j^*'s, are nearly evenly distributed when b_1 is small. That is, which is the time-threshold associated with 10% inventory sold, t_1^* is equal to about 0.10 (i.e., 10% of the sale period); t_2^*, which is the time-threshold associated with 20% inventory sold, is equal to about 0.20 (i.e., 20% of the sale period), and so on. The reason is that when the herding effect is insignificant, the customer's purchase propensity is almost constant with respect to the fraction of inventory sold. In such cases, the scarcity effect dominates, and disclosing inventory is desired only when the current fraction of inventory sold is larger than the fraction of sale period elapsed. However, as b_1 increases further, the herding effect becomes the dominating effect, and the time thresholds should increase, thus encouraging the platform to disclose inventory information.

Having discussed properties of the optimal thresholds, we now explain the behavior of R_3. Note that the fixed threshold policy can be viewed as a special case of the time-dependent threshold policy. That is, suppose j^* is the optimal fixed threshold such that the platform will disclose inventory information once the amount of product sold reaches j^*. Indeed, such a policy is equivalent to a time-dependent threshold policy with $t_j = 0$ for $j < j^*$ and $t_j = T$ for $j \geq j^*$.

For example, when $b_1 = 2$, from Figure 4.2 we have $j^* = 1$ (i.e., 10% of inventory sold) under the fixed threshold policy, and $t_j^* = T = 1$ for all $j \geq 1$ under the time-dependent threshold policy, making the two policies equivalent, and thus $R_3 = 0$ as indicated by Figure 4.1. In contrast, when $b_1 = 0.75$, then $j^* = 6$ (i.e., 60% of inventory sold) under the fixed threshold policy, which implies that the platform should disclose its inventory once 60% of its inventory is sold. Nevertheless, the resulting t_j^*'s from the time-dependent threshold policy are almost evenly distributed between 0 and T. In such cases, the optimal time-dependent threshold policy is different from the optimal fixed threshold policy, and the improvement of the former compared with the latter must be significant.

Fourth, since R_4 represents the improvement of the time-dependent threshold policy with respect to the better of the two simple policies, it has two peaks each corresponding to the respective peaks of R_2 and R_3. Moreover, the peak associated with R_2 occurs at a larger b_1 than that associated with R_3. This result implies that, when the herding effect is strong (b_1 is large), the improvement of switching from a simple policy to the fixed threshold policy dominates the additional improvement of switching from the fixed threshold policy to the time-dependent threshold policy, and vice versa.

Impact of other parameters pertaining to the observational learning effect. In addition to examining the impact of changing b_1, which is the most important parameter that governs the weight of the observational learning effect in the customers' purchasing decisions, we have also considered other parameters including ρ_{max}, which is the maximal weight associated with the quality signal in the updating of the perceived quality, and the distribution of the customers' prior beliefs of the product quality. We have conducted additional numerical experiments to examine the impacts of these parameters. However, to converse space, we refer readers to Appendix B.2 for more details about these extensions.

4.3.2 *Impact of the Scarcity Effect*

In our customer purchase propensity model, b_2 represents the weight associated with the scarcity effect. In addition, p_{so}^{ND}, p_{so}^{BD}, and $p_{so}(s, \tau)$ are the perceived sell-out probability by the customers under different scenarios, which are pertinent to the scarcity effect as well. In this subsection, we investigate impacts of these parameters on performances of different inventory disclosure policies.

Impact of the weight associated with the scarcity effect in customers' purchasing decisions. Figure 4.3 shows how performances of different policies change with b_2. Figure 4.4 depicts the optimal thresholds as b_2 varies. A higher value of b_2 means that the scarcity effect plays a more prominent role in customers' purchasing decisions.

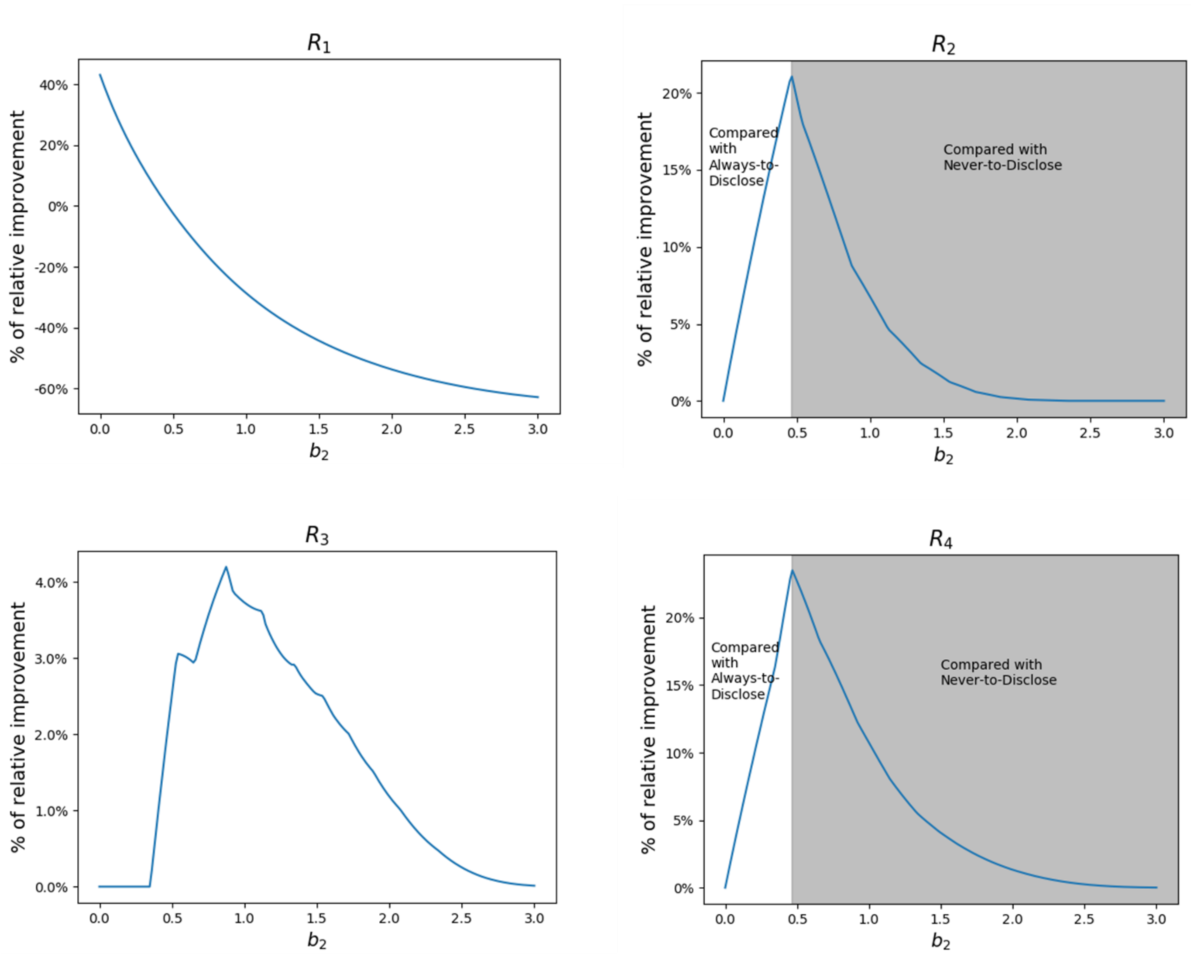

Figure 4.3. Measures of improvements as the weight attached to the scarcity effect, b_2, varies.

Note. $b_1 = 1.2$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

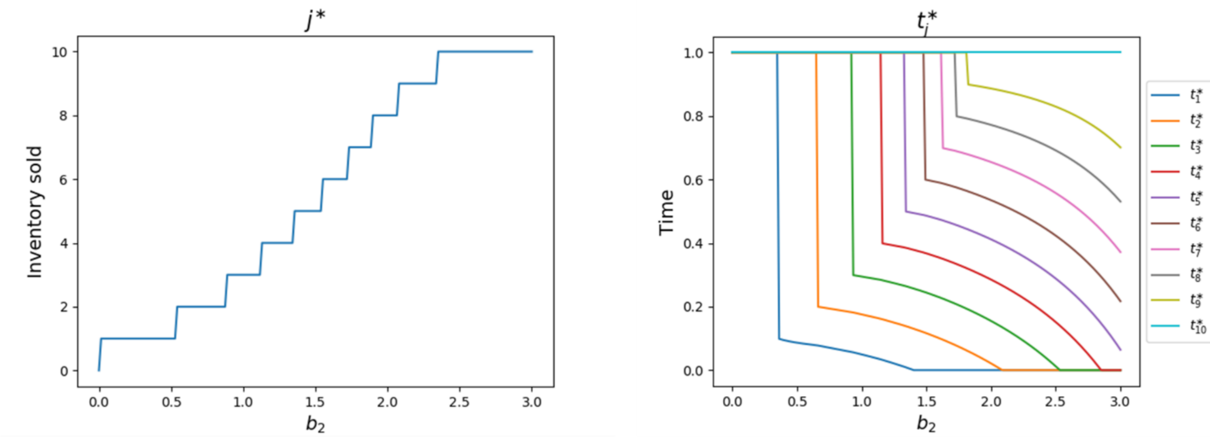

Figure 4.4. The optimal thresholds of the fixed (left) and time-dependent (right) threshold policies as b_2 varies.

Note. $b_1 = 1.2$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

First, consider R_1, the improvement of “always disclose” compared to “never disclose.” Under the former, as b_2 increases, the average demand rate is influenced more heavily by the scarcity effect. In particular, the negative consequence of slow sales becomes more serious. This worsens the performance of the “always disclose” policy and causes a decrease in R_1.

Second, we examine the improvement of the optimal fixed threshold policy, R_2. As can be seen from Figures 4.3 and 4.4, R_2 is unimodal with respect to b_2, and, interestingly, the optimal fixed threshold is increasing in b_2. The increasing trend of the optimal fixed threshold implies that, as b_2 increases, disclosing inventory becomes less favorable. The reason is as follows: The increase in b_2 amplifies the scarcity effect and makes the negative consequence of slow sales more serious, thus discouraging early disclosure of the inventory information. As for R_2, the existence of a peak is due to the fact that different simple policies perform better at different sides of the peak. Specifically, when b_2 is small, the “always disclose” policy dominates the “never disclose” policy. As b_2 increases, the fixed threshold policy gains more improvement as compared to the “always disclose” policy, as it alleviates the negative consequence of slow sales through raising the optimal fixed threshold for inventory disclosure. In contrast, when b_2 is moderate or large, the "never disclose" policy is the better policy of the two simple policies. In such cases, the advantage of the optimal fixed threshold policy compared to the "never disclose" policy diminishes as the scarcity effect amplifies.

Third, we investigate R_3 and the optimal time thresholds associated with different levels of inventory sold. Figure 4.3 shows that R_3 is first increasing and then decreasing in b_2. Moreover, as shown in Figure 4.4, the optimal time thresholds associated with different levels of inventory sold are decreasing in b_2. The decreasing trend of the time thresholds implies that disclosure of inventory information becomes less favorable, which is consistent with our reasoning above that a higher b_2 induces a more serious negative consequence of slow sales and thus discourages the platform to disclose its inventory. As for the impact of b_2 on R_3, it is similar to the impact of b_1 on R_3.

Fourth, the improvement of the optimal time-dependent threshold policy as compared to the two simple policies, R_4, can be viewed as a result of the combined effects of R_2 and R_3.

Impact of the customers' perceptions of the sell-out probability when inventory information is not available. Recall that p_{so}^{ND} denotes the customers' perceptions of the sell-out probability if inventory information will never be available. In contrast, if the customers know that the platform employs a threshold policy but so far the threshold for disclosure has not been reached, their estimate of the sell-out probability, p_{so}^{BD}, can even be smaller than p_{so}^{ND}, as the platform's not disclosing its inventory can be interpreted as a signal that there is still plenty of inventory available. Below, we examine impacts of the two parameters on different policies.

To begin with, we fix $p_{so}^{BD} = p_{so}^{ND}$ and vary p_{so}^{ND}. Figures 4.5 and 4.6 depict the impact of p_{so}^{ND} on performances of various policies and their respective optimal thresholds. The analogy between

Figure 4.3 (Figure 4.4) and Figure 4.5 (Figure 4.6) is worth noting, as it implies that the impacts of p_{so}^{ND} on the performance measures and the optimal thresholds are similar to those of the parameter b_2. Thus, we will not repeat the similar observations; instead, we provide a brief summary below.

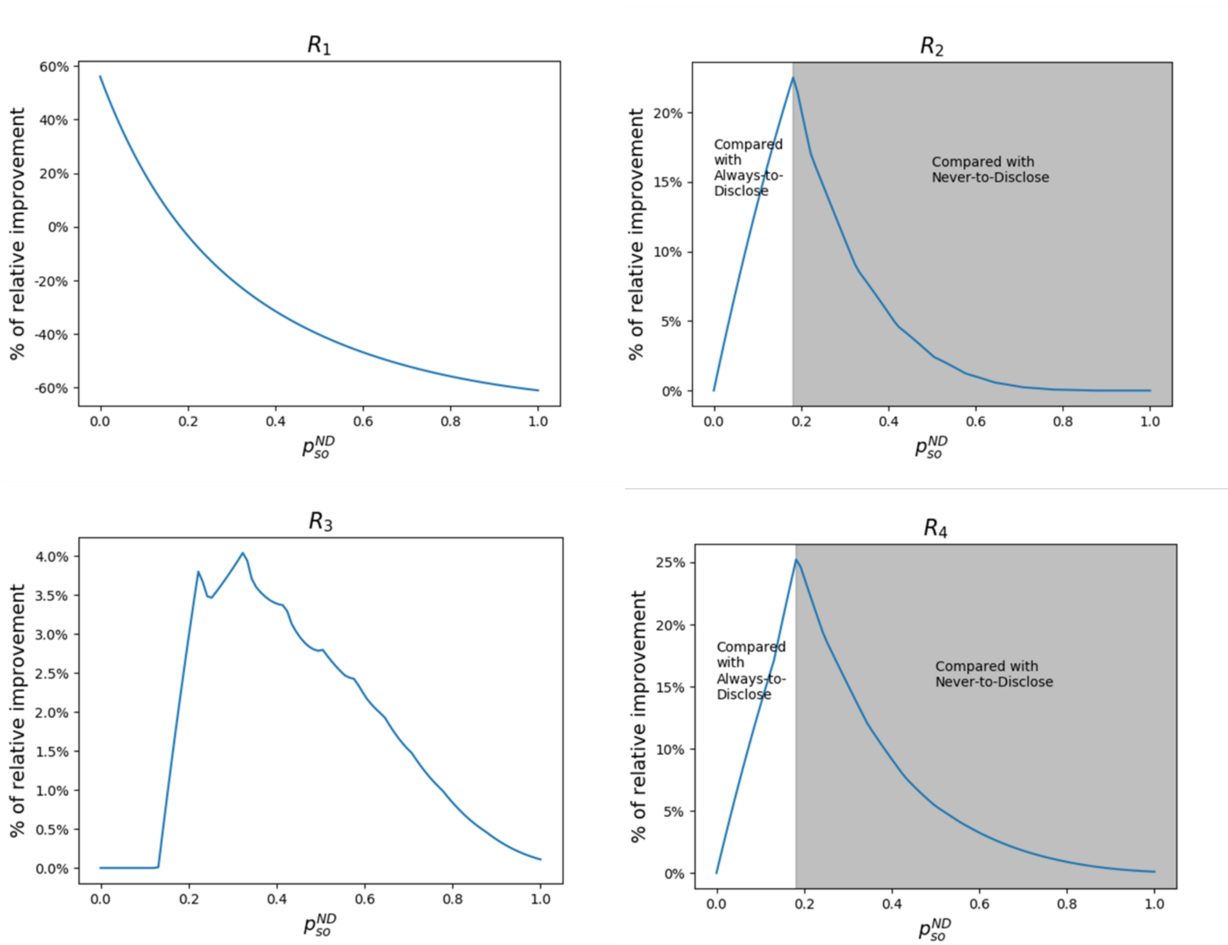

Figure 4.5. Measures of improvements as the customers' perceptions of the sell-out probability p_{so}^{ND} varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{BD} = p_{so}^{ND}$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

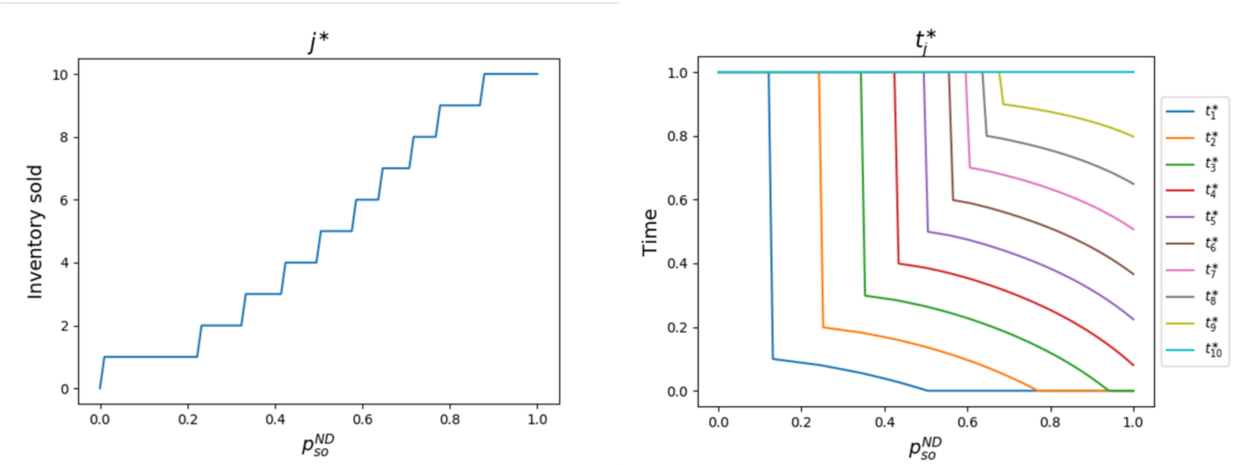

Figure 4.6. The optimal thresholds of the fixed (left) and time-dependent (right) threshold policies as p_{so}^{ND} varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{BD} = p_{so}^{ND}$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

First, higher values of p_{so}^{ND} are beneficial to the "never disclose" policy as customers are more likely to buy due to the scarcity effect, giving the "never disclose" policy an advantage over the "always disclose" policy. This explains the decreasing trend of R_1. Second, in the region where R_2 is increasing, the "always disclose" policy is better than the "never disclose" policy, while in the region where R_2 decreases, the opposite occurs. That is, as p_{so}^{ND} increases, the platform has less incentive to broadcast inventory information, which gives an advantage to the fixed threshold policy compared to the "always disclose" policy but a disadvantage compared to the "never disclose" policy. This explains the first increasing then decreasing trend of R_2. Third, R_3 exhibits a unimodal trend as p_{so}^{ND} increases. The reasoning behind this observation is the same along the line of reasoning used to explain the impact of b_2 on R_3. Moreover, as p_{so}^{ND} increases, the optimal fixed threshold increases while the optimal time thresholds associated with different levels of inventory sold decrease, which implies that an increase in p_{so}^{ND} discourages the platform to broadcast inventory information under both threshold policies. Last, R_4 is due to a combined effect of R_2 and R_3.

As a final remark, note that in our default setting, we have $p_{so}^{BD} = p_{so}^{ND}$. To address the concern that the customers may lower the estimate of the sell-out probability if they know the inventory information may be disclosed in the future but is not at the present time, we consider an extension to our numerical study with $p_{so}^{BD} \leq p_{so}^{ND}$. Specifically, we fix $p_{so}^{ND} = 0.3$ as in the default setting and vary the ratio of p_{so}^{BD} to p_{so}^{ND} between zero and one. Figure 4.7 depicts the four performance measures of various policies as the ratio of p_{so}^{BD} to p_{so}^{ND} varies.

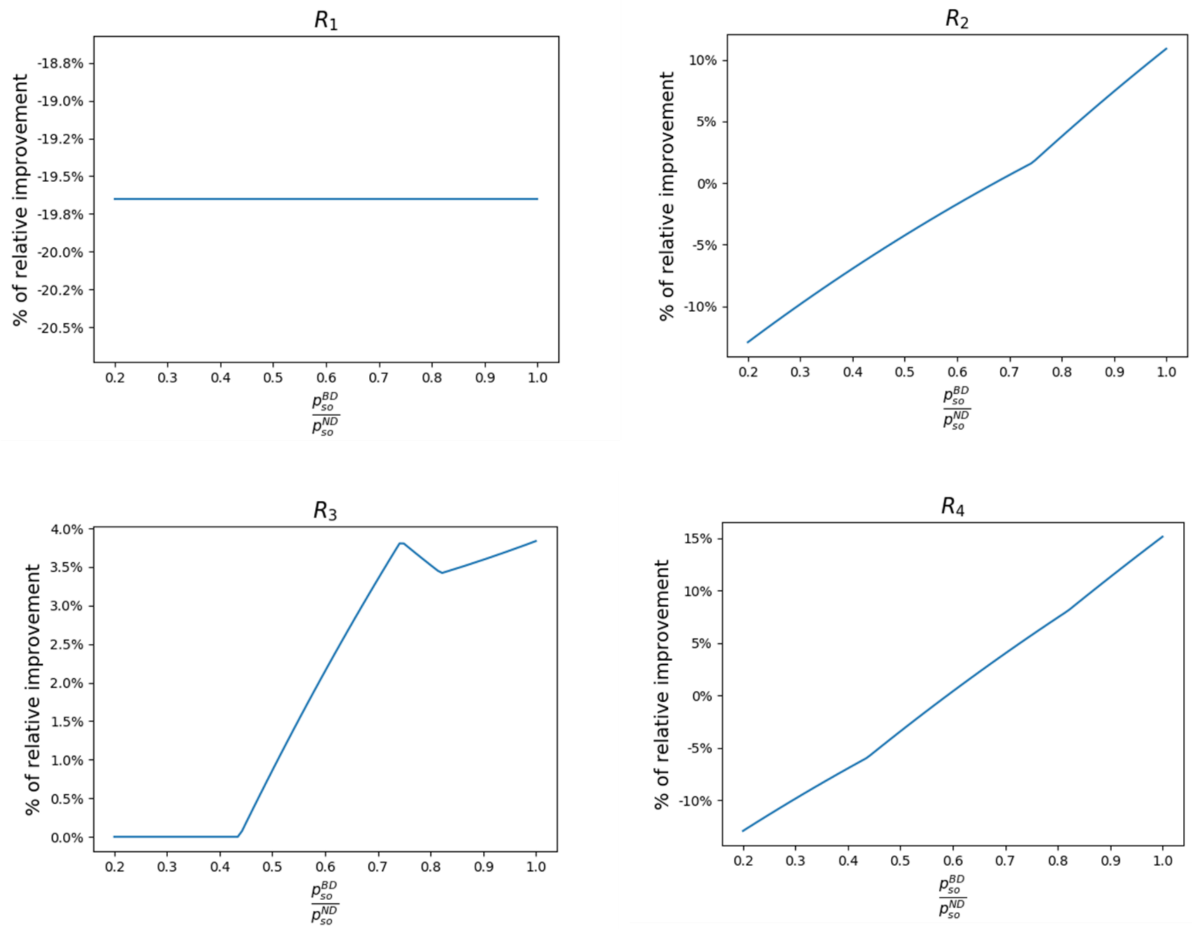

Figure 4.7. Measures of improvements as the ratio of p_{so}^{BD} to p_{so}^{ND} varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

As can be seen in Figure 4.7, R_1 stays constant as the performances of the "never disclose" and "always disclose" policies do not depend on p_{so}^{BD}. To be precise, under the "never disclose" policy, the customers' estimate of the sell-out probability is p_{so}^{ND}, which is no smaller than p_{so}^{BD}, whereas, under the "always disclose" policy, the customers' estimation of the sell-out probability will be updated in real time, $p_{so}(s,\tau)$, given their knowledge about the fraction of inventory sold s and the fraction of time elapsed τ. As for R_2, it is increasing with $(p_{so}^{BD}/p_{so}^{ND})$. Remarkably, when the ratio is too small, R_2 can even be negative, which implies that adopting the fixed threshold policy may backfire if the platform's not disclosing its inventory is interpreted as a signal of slow sales by the customers. In contrast, when the ratio grows larger, the negative impact of the customers' lower perceptions of the sell-out probability before disclosure diminishes, so the fixed threshold policy outperforms the simple policies, as expected. In general, R_3 also exhibits an increasing trend as the ratio $(p_{so}^{BD}/p_{so}^{ND})$ increases. With that said, the trend of R_3 is driven by the increase of p_{so}^{BD} alone, as the comparison between the fixed and time-dependent threshold policies are independent of p_{so}^{ND}. Last, R_4 can be viewed as a result of the combined effects of R_2 and R_3.

Table 4.4. Impact of different factors on the improvements of policies examined in this paper

	b_1	b_2	p_{so}^{ND}	p_{so}^{BD}/p_{so}^{ND}
In order to have higher R_1	large	small	small	N/A
In order to have higher R_2	medium to large	small to medium	small to medium	large
In order to have higher R_3	small to medium	small to medium	small to medium	large
In order to have higher R_4	medium to large	small to medium	small to medium	large

Finally, Tables 4.4 and 4.5 summarize impacts of different factors examined in this section on the performances of various policies and the associated optimal thresholds, respectively. We will discuss managerial insights derived from these tables in the conclusion.

Table 4.5. Impact of various factors on the optimal thresholds of the fixed and time-dependent threshold policies

With a higher value of	b_1	b_2	p_{so}^{ND}	p_{so}^{BD}/p_{so}^{ND}
Optimal fixed threshold	increases then decreases	increases	increases	Increases

Optimal time-thresholds associated with different levels of inventory sold	almost constant then increases	decreases	decreases	decreases

4.4 CONCLUSION

The extant operations management literature on managing online sales of limited inventory has provided empirical evidence for the impact of disclosing inventory availability information on the net sales. From the perspective of the platforms, they have to determine not only ***whether*** to broadcast its inventory information but ***when*** to broadcast it as well. To the best of our knowledge, this study is the first to analyze the timing of inventory information broadcast in the context of selling a product with a fixed initial inventory in a limited period, thus filling the gap between the extant empirical works and the platforms' needs to optimize their inventory disclosure policies.

Inspired by common inventory disclosure practices, we analyzed the following inventory disclosure policies: (i) the "*always disclose*" policy, (ii) the "*never disclose*" policy, and (iii) the *fixed threshold* policy. In particular, for the fixed threshold policy, we derived the expected total sales based on any given threshold. Furthermore, for cases where the average demand rate after the inventory disclosure is multiplicatively separable with respect to the amount of inventory sold and time elapsed, we obtained a closed-form expression for the expected total sales, making the search for the optimal threshold efficient. The fixed threshold policy, however, has a caveat that the disclosure is triggered once the inventory level drops below the fixed threshold, no matter ***when*** it occurs. To address the caveat, we proposed and analyzed a *time-dependent threshold* policy as follows. For any given inventory level at the present time, a time threshold should be calculated and compared to the current time. The platform should disclose its inventory information if and only if the current time is earlier than the time threshold associated with the current inventory level. The intuition behind the proposed policy is that, if the inventory level drops faster than expected, the platform should disclose its inventory information to send a signal to the customers that the product is popular and of high quality. Indeed, under two mild assumptions, we showed that the proposed policy is the optimal policy, and we devised a dynamic-programming-based algorithm to determine the optimal time thresholds associated with different levels of inventory sold.

In addition to devising procedures for policy optimization based on a general functional form of demand rates, we also developed a specific demand model based on customers' purchase propensity and the Bayesian updating framework, whereby customers' purchasing decisions are influenced by the *observational learning (herding)* and *scarcity* effects. The resulting functional form of the average demand rates is consistent with the empirical evidence in the extant literature. Through a comprehensive numerical study, we examined the relative performances of different policies. Here, we list the main managerial insights derived from those results.

- The "always disclose" policy outperforms the "never disclose" policy in cases where the observational learning effect dominates the scarcity effect in the customers' purchasing decisions. Those are cases where either the customers give a larger weight to the perceived quality than the perceived sell-out probability or the perceived sell-out probability is small.
- Under similar conditions (i.e., the observational learning effect dominates the scarcity effect), the optimal fixed and time-dependent threshold policies both can bring significant improvements compared to the better of the two simple policies. However, it is important to be aware of the possibility that employing a threshold policy may backfire. That is, if the customers know that the platform employs a threshold policy for its sales, they will interpret not disclosing at the present time as a signal of slow sales, resulting in lower demand. Hence, the platform should be mindful of such a negative consequence due to the threshold

policy, especially when the scarcity effect has a greater weight in the customers' purchasing decisions.

- Between the two threshold policies, the time-dependent threshold policy can bring a significant additional improvement compared to the fixed threshold policy given that the weight associated with the observational learning effect is relatively small, whereas, as noted above, the condition for the fixed threshold policy bringing a significant improvement compared to the two simple policies is that the weight associated with the observational learning effect is relatively large. In other words, when the learning effect dominates the scarcity effect, not only can the fixed threshold policy outperform the two simple policies, but the room for further improvement also is small, implying that the fixed threshold policy is near optimal. However, when both learning and scarcity effects are insignificant, there exists a large gap between the performances of the fixed and the time-dependent threshold policy, implying that it is desirable for the platform to employ the time-dependent threshold policy.

Finally, we would like to point out a few avenues for future research. One direction is to examine the optimal policy in the most general form. That is, although we assumed that the platform cannot "flip-flop" its inventory disclosure decision, it will be interesting to analyze a general policy without such a restriction. Another direction is to take into account the platform's pricing decisions during the sale period. For instance, one can examine the platform's joint decisions on pricing (fixed versus dynamic) and the disclosure of inventory information. Exploring those directions will lead to intriguing future research and valuable findings to complement this paper.

Chapter 5. A BLESSING OR A CURSE: THE IMPACT OF SAMPLING IN CROWDFUNDING MARKET

Crowdfunding has emerged as an essential source of funds for entrepreneurs in recent years. On crowdfunding platforms, investors can browse hundreds of projects in a short time, which reduces the search cost for investors and benefits fundraisers as well. In this paper, we focus on reward-based crowdfunding, where backers support founders in exchange for rewards (Hemer 2011). The market of reward-based crowdfunding has grown quite large with a promising future. The global transaction value of reward-based crowdfunding is projected to reach $1,082.9 million, with the total number of crowdfunding campaigns amounting to 184.0 thousand in 2021. It is predicted to increase with an annual growth rate of 2.62% in the next four years.

While the crowdfunding market is growing rapidly, it is still a young marketplace, and much is needed to be improved. Of all aspects, information asymmetry is an important issue, and it has existed since the birth of crowdfunding (Agrawal et al. 2014). While founders hold private information about their projects, backers can only receive incomplete and usually carefully selected information of projects from founders on the crowdfunding platform. Compared to equity-based crowdfunding markets, which are usually regulated by authorities and the government, reward-based crowdfunding markets lack strict information disclosure regulations (Cascino et al. 2019). Therefore, information asymmetry can cause severe adverse selection problems on our focused platforms and lead to market inefficiencies (Akerlof 1970). One effective way for backers to learn project quality is observational learning, or herding (Zhang and Liu 2012). Through rational herding, backers are more attracted by projects with a higher amount raised, which indicates that the project is of better quality. Meanwhile, crowdfunding platforms also design various policies to address the information asymmetry problem. Research focusing on platforms' information control policies (Burtch et al. 2016; Zhou et al. 2021) finds the benefit of revealing more information of both projects and previous backers. Some platforms leverage prefunding, a novel feature that allows founders to share project information with potential backers before the fundraising begins, to help backers learn more about the project (Wei et al. 2021).

Our research aims at an innovative design that has yet drawn little attention in the crowdfunding market – product sampling. Product sampling is a promotional marketing strategy that can largely reduce customers' uncertainty of product quality (Heiman et al. 2001) and thus boost product sales in both offline and online markets (Schultz et al. 1998; Chellappa and Shivendu 2005; Lin et al. 2019; Liu et al. 2021). However, in the context of crowdfunding, the role of product sampling is more sophisticated. In the sampling campaign, founders offer free samples, which are usually one of the numerous reward-tiers of the project, for backers to apply for. Lucky users that receive the free samples need to provide feedback reports after usage. In this process, product sampling alleviates the information asymmetry problem through two channels: (1) emitting signals regarding project quality upon joining the sampling campaign, and (2) the feedback reports which provide evaluations of the sampling product from successful applicants. The first channel resembles the signaling effect on e-commerce platforms, where e-tailers provide free samples to convince customers that their products are of high quality (Liu et al. 2021). The second channel is particularly interesting and serves as the critical difference between sampling campaigns on crowdfunding platforms and those on e-commerce platforms. Different from e-commerce platforms where product reviews from previous consumers are always available, without sampling campaigns, backers cannot take advantage of reports or ratings from their peer backers on the crowdfunding project to reduce quality uncertainty of the project. Now, with the help of feedback reports written by peers, backers obtain valuable and often more detailed information on the project's quality. As the information asymmetry problem can be substantially addressed by feedback reports, backers may highly rely on these reports to evaluate the project, and their behaviors regarding the sampling campaigns are worth investigating.

The two channels through which product sampling alleviates information asymmetry may affect the fundraising process in opposite directions. On the one hand, taking part in sampling campaigns can signal high product quality, as low-quality projects are not confident enough about their products' quality and will not risk sampling their products. The signaling effect of the sampling campaign can benefit founders. However, on the other hand, product sampling can also act as a double-edged sword. As feedback reports are of high value for backers, backers may strategically delay making decisions of whether to support the project or not before feedback reports are published. This is a unique mechanism of product sampling in our context. Previous research indicates that customers may delay making decisions for many reasons, such as relying on advice from others, expecting the price to fall, and anticipating quality improvement (Greenleaf and Lehmann 1995). This phenomenon has attracted many researchers in the field of operations management to study the dynamic pricing strategy with forward-looking customers who may delay their purchase (Levin et al. 2009; Yu et al. 2016; Papanastasiou and Savva 2017), and it could become a serious problem for founders if they are unaware of it. Besides backers' strategic delay of decision-making, another potential issue of product sampling is that more information revealed to backers does not guarantee a surge of demand (Zhang and Liu 2012). In short words, product sampling may delay demand and even hurt demand after backers get more information from feedback reports.

Finally, besides the mechanisms alleviating information asymmetry problems, we need to consider another merit of participating in sampling campaigns, which is to draw attention from more backers. As backers cannot browse all projects on the platform, it is important for fundraisers to make their projects noticed by as many backers as possible. The platform usually creates a sub-platform for sampling projects, and it is a great opportunity for founders to make their projects known by more backers. As many factors should be considered in the sampling campaign on crowdfunding platforms, it is essential to analyze all mechanisms behind it and find the true impact of sampling campaigns on projects. This research fills the research gap in crowdfunding and product sampling literature, and it also provides managerial implications to both the founders and the platforms.

In this paper, we perform an empirical analysis of product sampling on crowdfunding platforms. We study both sides of the market and expect to find out how free product sampling affects projects' fundraising process. Specifically, our first research question is: (1) What is the effect of product sampling on crowdfunding platforms? Also, as there is empirical evidence that the number of samples and the report score (or online review in the context of e-commerce platforms) serve as effective quality signals, we want to know that: (2) Do the number of samples and feedback reports' score moderate the impact of sampling campaigns on crowdfunding platforms?

To answer these questions, we leverage a rich and unique dataset from JD crowdfunding, one of the largest crowdfunding platforms in China. JD crowdfunding created a sub-platform for free product sampling. Founders can participate in the sampling campaign by offering free samples on the platform for backers to apply. Users chosen to receive free samples need to provide feedback reports on the platform after using the product. We construct a structural model to estimate the effect of the sampling campaign on projects. We use a BLP-style aggregate demand model (Berry et al. 1995) with a consideration set model (Goeree 2008), where the backer is assumed to only notice projects in her consideration set. We model both the demand and supply sides of the market to eliminate the selection bias.

We find evidence of positive but decreasing effects on backers' attention to projects in all three stages of the sampling campaign in chronological order, namely Application Stage, Trial Stage, and Report Stage. However, joining the sampling campaign lowers backers' utility towards supporting the project in the latter two stages. We also find positive signaling effects of the number of samples provided and the feedback report's score in ReportStage. Further counterfactual analysis shows that the demand for most projects increases in Application Stage and decreases in Trial Stage. In Report Stage, about half of the projects' demand increases, while the other half decreases.

Our research makes considerable contributions to academia. First, we fill the research gap in crowdfunding literature by investigating the effect of product sampling, a novel design of campaigns on crowdfunding platforms, and analyzing the interesting mechanisms behind it. Previous literature studies the impacts of various mechanisms and policies used to solve the information asymmetry issue on crowdfunding platforms (Zhang and Liu 2012; Burtch et al. 2016). Our study adds to this body of literature by documenting the performance of product sampling. We break down the influence of this innovative design into drawing more attention, signaling better quality, and the probably unexpected strategic delays of backers' decision-making to better understand the true impact of product sampling on crowdfunding platforms. We also provide meaningful managerial implications for both fundraisers and the crowdfunding platform when performing product sampling campaigns.

Second, we contribute to the product sampling literature by studying the impact of sampling in the context of the crowdfunding market. Product sampling has been studied in both online and offline markets (Heiman et al. 2001; Cheng and Liu 2012; Lin et al. 2019), where it is proved to be an effective promotional strategy. However, in our study, we point out that product sampling may not be desirable for founders to boost demand. The extraordinary value of feedback reports from users of free products distorts the effect of product sampling in the crowdfunding market. Our study provides empirical evidence of the impact of sampling design on crowdfunding platforms.

Our paper also extends the literature on the signaling effect by identifying the number of samples and the score of feedback reports as valid signals in the crowdfunding market. Ratings and reviews are typical signals on online platforms such as Amazon and Yelp. The number of samples is found to be an effective quality signal of products on e-commerce platforms (Liu et al. 2021). Our results show that they also serve as signals of product quality on crowdfunding platforms. Our work also makes methodological contributions to the IS literature. To the best of our knowledge, this study is the first one using the consideration set model in the IS literature to account for the limited attention of backers or customers on online platforms. Future work can leverage our model when a decision-maker faces so many options that she cannot browse all of them.

5.1 DATA

Our data set comes from JD crowdfunding, a large online crowdfunding platform in China. JD crowdfunding was launched in 2014 by JD, one of China's most influential e-commerce companies. In 2019, over 2,000 projects succeeded on the JD crowdfunding platform, with over ¥250 million raised from backers.

JD crowdfunding resembles large reward-based crowdfunding platforms in the U.S., such as Kickstarter and Indiegogo. On the JD crowdfunding platform, entrepreneurs with innovative ideas but are lack of enough funds can establish a crowdfunding project to attract potential backers. On the project's webpage, the fundraiser needs to provide the title, objective amount, and due date of the project together with the project description, which can include text, pictures, and videos. Also, as JD crowdfunding is a reward-based crowdfunding platform, the founder should offer different tiers of rewards with detailed descriptions for backers to choose from. Once the project starts to raise money, the real-time raised amount will also appear on the webpage of the project. If the project succeeds, backers will receive corresponding rewards. If the project fails, backers will receive a refund. Figure 5.1 shows an example of a project in progress.

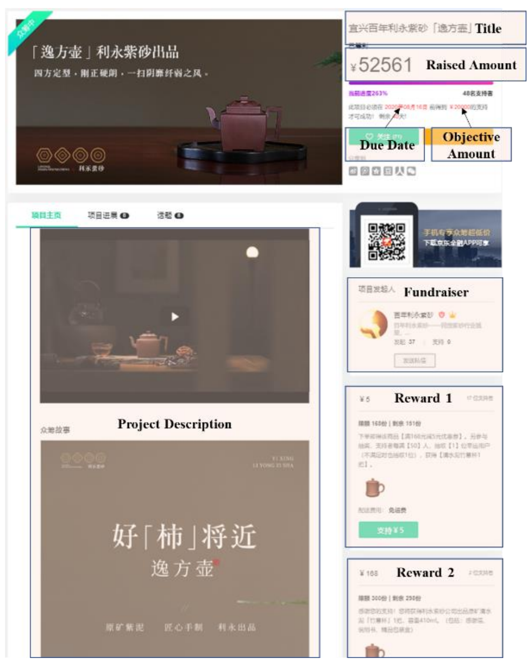

Figure 5.1. Homepage of a Project

The free product sampling platform, sampling platform hereafter, of JD crowdfunding launches in 2016 as an auxiliary platform of JD crowdfunding. Fundraisers interested in joining the sampling campaign can contact the staff of JD crowdfunding to offer free products on the sampling platform. An example of the webpage of a sampling project is shown in Figure 5.2. Backers can observe the name and the value of the sampling product, the stage of the sampling campaign, number of samples, number of applicants, a link to the webpage of the crowdfunding project, and feedback reports (if available) on the sampling webpage. Though the webpage shows that there are four potential stages of the project, we actually observe no projects in the first stage (pre-application), and the last stage (post-trial) does not coincide with the release dates of feedback reports or the due date of the project. We thus decide to redefine stages of the sampling campaign as three stages: Application Stage, Trial Stage, and Report Stage. In Application Stage, backers can apply for free samples. Then in Trial Stage, lucky users selected by the platform receive free samples and try them. Finally, in Report Stage, users provide feedback reports on the sampling platform, which can also be observed on the webpage of the crowdfunding project. As users of a specific sampling campaign provide reports on different dates, we define the start date of the Report Stage to be the date that the first report on the sampling product is posted. The timeline of a crowdfunding project participating in a sampling campaign can be described in Figure 5.3.

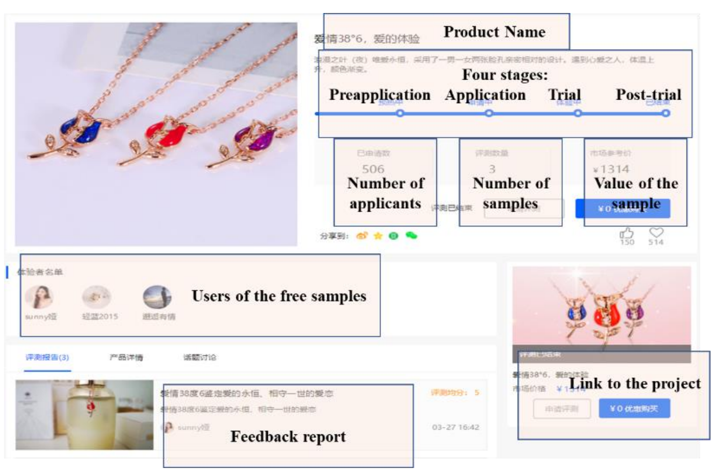

Figure 5.2. Webpage of a sampling product

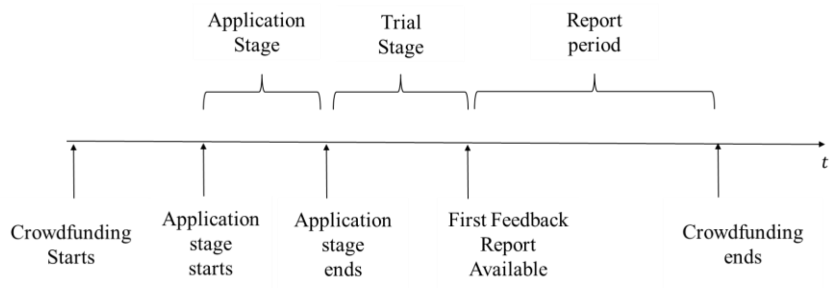

Figure 5.3. Timeline of a sampling project

We crawl the data of crowdfunding projects and sampling campaigns from June 2017 to January 2019 on a daily basis. For projects, we collect both project information and information on the reward. The panel data we collect consists of 1,362,200 records on the project-reward level. Out of all 7,209 projects in our dataset, 533 participated in the sampling campaign. We show the summary statistics of project-reward attributes in Table 5.1. In Table 5.1, Success indicates if the project is successful. Delivery Time is the time within which the project promises to deliver corresponding rewards to backers if they succeed. Sampling attributes in Table 5.1 show the statistics of 533 projects joining the sampling campaign.

Table 5.1. Summary Statistics

Variable	Mean	Std. Dev.	Min	Median	Max
Project Attributes					
Objective Amount	103,025.48	208,084.36	100	100,000	10,000,000
Median Price	1,359.45	7,923.37	1	400	527,950
Duration	35.64	10.59	1	31	113
Success	0.74	0.44	0	1	1
Delivery Time	20.83	13.72	0	20	300
Number of Rewards	6.80	3.99	1	6	108
Reward Attributes					
Price	4,513.61	37,344.85	1	399	5,000,000
Number of Backers	113.69	653.58	0	10	53,336
Sampling Attributes					
Number of samples	5.77	4.54	1	5	32
Number of Applicants	884.74	1,034.89	2	580	13,107
Sample Price	1,721.63	9,289.10	36	599	193,800
Number of Reports	4.73	3.50	1	4	26
Report Score	4.78	0.26	3.66	4.85	5

5.2 MODEL

We consider both sides of the crowdfunding market in our model. At the demand side, we combine a consideration set model, which describes the fact that backers may not notice all projects on the platform, and a BLP-style aggregate demand model to characterize the behavior of backers. The effect of the sampling campaign can be broken down into its impacts on backers' attention to the project and their utility of supporting the project. The supply side of the market is not only added to analyze the decision process of founders on sampling campaigns but also necessary to account for the potential selection bias at demand side of the market. For example, founders choosing to offer free samples may be systematically different from those not joining the sampling campaign. To solve this endogeneity issue, we construct a structural model that correlates the fundraiser's decision process and the utility that backers gain from investing in the project.

5.2.1 *Model of Backers*

On the demand side, we treat online crowdfunding in China as a marketplace. We aggregate our daily-level data to weekly-level data to make it more consistent with the concept of a market. We construct a discrete choice model with consideration sets for backers and follow the approaches in Berry et al. (1995) (BLP) and Goeree (2008) to approximate market shares of project-rewards in the marketplace using aggregated demand of backers. We first introduce the utility function for backers, then construct the consideration set model for backers.

We use a multinomial logit model as the discrete choice model for the backer. The utility for customer i on project-reward j at time t (at weekly level) is defined as:

$$u_{ijt} = x_{jt}^T \beta + \mu_t + \xi_{jt} + \varepsilon_{ijt} \tag{5.1}$$

In the above equation, x_{jt} includes attributes of project-reward j at time t. Attributes in x_{jt} are variables that backers can easily find on the project's webpage to help them evaluate the project. Specifically, in our basic model, x_{jt} includes indicators of different stages of the sampling campaign, namely *ApplicationStage*, *TrialStage*, *ReportStage*, and other variables, including *Price, log(ObjectiveAmount)*, *Duration*, *DeliveryTime*, *HasDeliveryTime, Category*, *MedianPrice, DaysElapsed*, *ClaimRatio*, *FullRatio*, *RaisedRatio*, *Success*, and a constant term. As mentioned above in Section 5.1, we redefine three stages of free product sampling, and we use three dummy variables to investigate their effects on backers' utility separately. We take the logarithm transform of *Objective Amount* as it is extremely right-skewed. *Category* are dummy variables indicating which category the project belongs to. *MedianPrice* is the median price of all reward options in the project. We add this variable to control for price at the project level and absorb the potential effect of product quality to alleviate the endogeneity problem. *HasDeliveryTime* indicates that whether the project shows the delivery time of their products or not. *ClaimRatio* is the proportion of the project-reward *j* already claimed. It can control for peer influence and scarcity effect at the reward level. *FullRatio* is the proportion of rewards in the project that is already full and is added to control any spillover effect from full reward-tier to other reward tiers. *RaisedRatio* is the percentage of money received compared to the objective amount. *Success* indicates whether the project has already reached its objective amount or not.

In our heterogeneous model, we add interaction terms between dummy variables of each stage and the moderators, namely *log(SampleNum)* and *ReportScore*. We take the logarithm transform of *SampleNum* after adding one to it, as its distribution is right-skewed, as shown in Table 5.1. We calculate *ReportScore* as the average score of all feedback reports of the focal project. Specifically, we add *ApplicationStage * log(SampleNum)*, *TrialStage * log(SampleNum)*, *ReportStage ** *log(SampleNum)*, and *ReportStage* * *ReportScore* to capture the moderating effects of the number of samples on all three stages and the moderating effect of feedback report's score on the last stage.

The μ_t in Equation (5.1) indicates time dummies, and we use month dummies in our model. ξ_{jt} is the mean utility error of project-reward j at time t, and ε_{ijt} is the type-I extreme value distributed idiosyncratic error. We can further divide ξ_{jt} into two parts: a project-level random effect $\zeta_{p(j)}$ and an error term ψ_{jt}. Here $p(j)$ denotes the project that the project-reward j belongs to. Different rewards within the same project share the same random effect $\zeta_{p(j)}$. The project-level random effect represents unobservable attributes that affect backers' utility, such as the quality of the project. To simplify expressions, we define the mean utility of the project-reward j at time t to be:

$$\delta_{jt} = x_{jt}^T\beta + \mu_t + \zeta_{p(j)} + \psi_{jt} \tag{5.2}$$

We next introduce the consideration set model. As JD crowdfunding is a large platform, it is almost impossible for backers to be aware of all projects ongoing at the same time. Hence, one would consider it reasonable to assume that the backer forms her own consideration set of projects before making the decision to support. Following the consideration set model in Goeree (2008), we define the probability that project p catches the attention of a backer at time t as:

$$\varphi_{pt} = \frac{1}{1 + \exp\left(-z_{pt}^T\alpha\right)} \tag{5.3}$$

In this equation, z_{pt} includes attributes that may affect the probability of the project to be noticed. To determine which variables to include in z_{pt}, we refer to the sorting method of project listings on JD crowdfunding which can largely affect the probability of the projects to be noticed. In our basic model, the variables in z_{pt} are: *ApplicationStage*, *TrialStage*, *ReportStage*, *DaysElapsed*, *Duration*, *log(Raised)*, *log(Backers)*, *Category*, and a constant term. As the platform creates another webpage for projects with free sampling, we assume that free sampling also affects the probability of the project to be considered. *Raised* is the total amount raised, and *Backers* is the total number of backers for the project. We add one before taking logarithms for these two variables. In our heterogeneous model, we add the same interaction terms as in the backer's utility model to investigate the moderating effects of the number of samples and feedback report's score on the project's probability of being noticed.

With our consideration set model, the backer's investment behavior is characterized as follows. The potential backer first notices some projects with probabilities defined in Equation (5.3) and puts the projects she is aware of into her consideration set. She then forms utilities on supporting these projects following Equation (5.1). And finally, she picks the one with the highest utility out as her final choice. Thus, we can derive market shares of project-rewards based on the two models we built above. Normalizing the mean utility of the outside option to be 0 and assuming that all backers are aware of the outside option, the market share of project-reward j at time t can be found by:

$$s_{jt} = \sum_{\mathcal{S} \in \mathcal{C}_{p(j)}} \prod_{l \in \mathcal{S}} \varphi_{lt} \prod_{k \notin \mathcal{S}} (1 - \varphi_{kt}) \frac{\exp(\delta_{jt})}{1 + \sum_{p(r) \in \mathcal{S}} \exp(\delta_{rt})} \tag{5.4}$$

Here $\mathcal{C}_{p(j)}$ is the set of all consideration sets that include project $p(j)$. For every consideration set $\mathcal{S} \in \mathcal{C}_{p(j)}$, we calculate the probability of forming this set and the conditional market share of project-reward j within this set. Aggregating conditional market shares from all $\mathcal{S} \in \mathcal{C}_{p(j)}$, we can get the market share s_{jt}. As we cannot directly inverse Equation (5.4) to find an expression of δ_{jt} using market shares, we need to use the contraction mapping method in BLP to derive correct mean utilities δ_{jt} from real market shares S_{jt} in data. As calculating market shares from Equation (5.4) is intractable, we simulate consideration sets for backers. Details of the estimation method can be found in Section 5.2.3. We obtain the information of the market share of the outside option on Zhongchoujia, an online platform providing analysis and news of crowdfunding in China, to help us determine the real market shares.

5.2.2 *Model of Fundraisers*

At the supply side of the crowdfunding market, we characterize the decision process of fundraisers on how many free samples to provide using a log-linear model:

$$\log\left(SampleNum_p\right) = w_p^T \gamma + \omega_p \tag{5.5}$$

In Equation (5.5), the dependent variable is the logarithm of the number of free samples provided by project p. We add one to it before taking the logarithm. If the project does not participate in sampling campaigns, its $SampleNum$ is set as 0. w_p represents attributes of the project p that can affect the founder's decision on how many free samples to offer. Specifically, w_p includes *MedianPrice*, *log(ObjectiveAmount)*, *Duration*, *DeliveryTime*, *Category*, *SampleProp*, and a constant term. *SampleProp* is the percentage of projects participating in the sampling campaign at the start date of project p.

ω_p is the error term in the decision process in Equation (5.5). It indicates the factor that affects the founder's decision but is unobservable to researchers. As the founder with a high-quality project may be confident about their products and become more interested in joining the sampling campaign or providing more free samples, we need to eliminate this potential selection bias. To be specific, we assume that the project level random effect ζ_p (we drop the dependency on project-reward j for simplicity) and the error term ω_p follow a bivariate normal distribution:

$$\begin{pmatrix} \zeta_p \\ \omega_p \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\zeta^2 & \rho\sigma_\zeta\sigma_\omega \\ \rho\sigma_\zeta\sigma_\omega & \sigma_\omega^2 \end{pmatrix} \right) \tag{5.6}$$

5.2.3 *Estimation*

We use simulated maximum likelihood estimation (simulated MLE) to jointly estimate all parameters. We first explain how to calculate simulated market shares s_{jt} using mean utilities δ_{jt} in Equation (5.4), and how to derive mean utilities from real market shares S_{jt}. We sample N_s backers at every time t. We calculate the probability to be noticed for every project φ_{pt}, and we simulate the consideration set of every backer using these probabilities. We can then find the conditional market shares of project-rewards for every backer. Finally, we aggregate the conditional market shares from backers by taking the mean of all simulated conditional market shares to find simulated s_{jt}. With the simulation method above, we can apply the contraction mapping method in BLP to calculate mean utilities δ_{jt} from real market shares S_{jt}. Details about the contraction mapping method can be found in Berry et al. (1995).

After deriving mean utilities from real market shares, we can calculate the likelihood of a project p as follows:

$$L_p(\xi_{jt}, \omega_p) = f_\omega(\omega_p) \int_{-\infty}^{+\infty} \prod_{j \epsilon \mathcal{J}(p)} f_\psi(\xi_{jt} - \zeta_p)\, f_{\zeta|\omega}(\zeta_p|\omega_p) d\zeta_p \tag{5.7}$$

Here $f_\omega(\cdot)$ is the probability density function of ω_p, $f_\psi(\cdot)$ is the probability density function of ψ_{jt} (we assume it to be normal as well), and $f_{\zeta|\omega}(\cdot)$ is the probability density function of ζ_p conditional on ω_p. We use $\mathcal{J}(p)$ to denote the set of project-rewards j that belongs to the project p. We express the likelihood as a function of ξ_{jt} and ω_p, as we can find their values with mean utilities calculated from simulated consideration sets and supply-side parameters. The idea of the likelihood function is to first find out the likelihood from the supply side of the market (i.e., $f_\omega(\omega_p)$) then find out the likelihood from the demand side of the market conditional on ω_p. As all random variables follow normal distributions, we can obtain an analytical likelihood function by integrating out the part in the integral, and we do not need simulations to calculate the likelihood at this step.

5.3 RESULTS

The results of our models are presented in Table 5.2. We normalize *Price*, *MedianPrice*, *Duration*, and *DeliveryTime* in utility and fundraiser models to express estimated parameters on a better scale. We also normalize DaysElapsed, Duration, log(Raised), log(Backers) in the consideration set model for the same reason. We first estimate a naïve model which only considers the demand side (Equation (5.1)) of the market without taking consideration set into account. In the naïve model, we keep the assumption that project-rewards within a project share the same random effect. The corresponding result is shown in column (1) of Table 5.2. We find that *ApplicationStage* raises backers' utilities, but the other two stages exert no significant effects on backers' utilities. However, without proper control from the supply side of the market and characterizing backers' awareness of projects, the result can be biased. Thus, we turn to the result of our basic model.

Table 5.2. Effect of Free Product Sampling on Crowdfunding Projects

	(1) Naïve Model		(2) Basic Model		(3) Heterogeneous Model	
Utility Model:						
ApplicationStage	0.209***	(0.038)	-0.018	(0.044)	-0.207*	(0.121)
*ApplicationStage * log(SampleNum)*	/		/		0.116*	(0.067)
TrialStage	0.004	(0.039)	-0.174***	(0.045)	-0.143	(0.124)
*TrialStage * log(SampleNum)*	/		/		-0.013	(0.069)
ReportStage	0.057	(0.038)	-0.078*	(0.045)	-2.703***	(0.488)
*ReportStage * log(SampleNum)*	/		/		0.246***	(0.067)
*ReportStage * ReportScore*	/		/		0.456***	(0.098)
Price	-0.142***	(0.004)	-0.141***	(0.004)	-0.141***	(0.004)
MedianPrice	-0.009	(0.010)	-0.007	(0.010)	-0.006	(0.009)
log(ObjectiveAmount)	0.233***	(0.017)	0.235***	(0.017)	0.235***	(0.017)
Duration	0.014***	(0.001)	0.024***	(0.001)	0.024***	(0.001)
DaysElapsed	-0.024***	(0.000)	-0.032***	(0.001)	-0.032***	(0.000)
DeliveryTime	-0.004***	(0.001)	-0.004***	(0.001)	-0.004***	(0.001)
HasDeliveryTime	0.110	(0.106)	0.119	(0.104)	0.120	(0.104)
ClaimRatio	0.807***	(0.016)	0.820***	(0.016)	0.820***	(0.016)
RaisedRatio	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
FullRatio	0.187**	(0.080)	0.226***	(0.080)	0.227***	(0.079)
Success	-0.211***	(0.012)	-0.175***	(0.012)	-0.175***	(0.012)
CategoryDummies	√		√		√	

MonthDummies	√	√	√
Constant	-13.099*** (0.224)	-13.208*** (0.219)	-13.207*** (0.219)
σ_ζ	4.031*** (0.009)	0.797*** (0.009)	0.796*** (0.009)
σ_ψ	7.831*** (0.003)	1.389*** (0.003)	1.389*** (0.003)
Consideration Set Model:			
ApplicationStage	/	0.781*** (0.036)	0.705*** (0.019)
*ApplicationStage * log(SampleNum)*	/	/	0.038*** (0.002)
TrialStage	/	0.482*** (0.018)	0.230*** (0.010)
*TrialStage * log(SampleNum)*	/	/	0.107*** (0.005)
ReportStage	/	0.347*** (0.009)	0.183*** (0.007)
*ReportStage * log(SampleNum)*	/	/	0.043*** (0.002)
*ReportStage * ReportScore*	/	/	0.021*** (0.001)
DaysElapsed	/	0.044*** (0.000)	0.045*** (0.000)
Duration	/	-0.056*** (0.000)	-0.056*** (0.000)
log(Raised)	/	-0.177*** (0.001)	-0.177*** (0.001)
log(Backers)	/	-0.303*** (0.002)	0.303*** (0.001)
CategoryDummies	/	√	√
Constant	/	3.144*** (0.010)	3.144*** (0.013)
Founder Model:			
MedianPrice	/	-0.021*** (0.006)	-0.021*** (0.006)
log(ObjectiveAmount)	/	0.036*** (0.009)	0.036*** (0.009)
Duration	/	0.000 (0.001)	0.000 (0.001)
DeliveryTime	/	0.000 (0.000)	0.000 (0.000)
HasDeliveryTime	/	0.034 (0.057)	0.033 (0.057)
SampleProp	/	1.916*** (0.239)	1.912*** (0.239)
CategoryDummies	/	√	√
Constant	/	-0.489*** (0.115)	-0.486*** (0.115)
σ_ω	/	0.474*** (0.004)	0.474*** (0.004)
ρ	/	0.098*** (0.016)	0.089*** (0.017)
Number of observations	137,104	137,104	137,104
AIC	/	502,195.90	502,169.37

*Note: Standard errors in parentheses; Significant levels: *** p <0.01, ** p <0.05, * p <0.1*

The result of our basic model is in column (2) of Table 5.2. The impact of the sampling campaign is two-fold. First, we find that the sampling campaign raises the project's probability of being noticed. As the platform shows projects with free sampling products on a separate page and can also make sampling projects rank higher in the regular list of projects, sampling projects benefit from sampling by drawing more attention from backers. This advantage peaks at *Application Stage* then decreases in *Trial Stage* and *Report Stage* as the attention of backers fades away. However, on the other hand, backers seem to obtain less utility from project-rewards of sampling projects in the latter two stages. *Trial Stage* harms the project the most, while the effect of *Application Stage* is the least and insignificant. The explanation of our astonishing findings is as follows. In crowdfunding markets, backers are almost impossible to get evaluations of the project's quality from peer backers. This is different from e-commerce platforms, where other customers' reviews and reports on products can be easily found. Thus, backers cherish the opportunity of acquiring information on project quality from peer backers so much that they would rather wait for feedback reports to help them determine whether to support the project or not. In *Application Stage*, the project just joins the sampling campaign, and it can signal its high quality as a brand-new project in a short time. This effect neutralizes the effect of backers' strategic delay, so the aggregated effect on utility is slightly negative and insignificant. However, in *Trial Stage*, the application is over, and the project may not be able to signal high quality as much as in the previous stage. To make things worse, backers know that feedback reports are approaching, and they are more willing to wait for evaluations than to invest in the project now. These two factors contribute to the sharp decrease of backers' utilities in *Trial Stage*. In *Report Stage*, the treatment effect is still negative and significant. The potential explanation is that backers overestimated the quality of the project in previous stages, and the feedback reports show that the project cannot meet their expectations. In short, the sampling campaign is a double-edged sword for founders. It can help the project attract the attention of more backers, but it also harms the project by decreasing the utilities of backers on supporting the project.

In addition, the correlation coefficient is positive and significant, which indicates that the founder of a project with higher quality (so it is more appealing to backers) is more likely to participate in the sampling campaign and offer more sampling products. This phenomenon coincides with our expectation, as founders of better projects are more confident with their products and are willing to release more project information by offering free samples to backers.

Without the model of the supply side, we can only observe an upward biased estimation of treatment effects on backers' utility in the naïve model, which leads to incorrect conclusions.

In column (3), we show the moderating effect of two signals of project quality, the number of samples provided, and the feedback report's score. We first find that projects with more samples and higher report scores can attract more backers. The reason is probably that the crowdfunding platform ranks projects higher if they provide more samples or their report scores are higher. The platform's behavior can be explained as follows. As a crowdfunding platform, it also knows that more samples and higher scores are positive quality signals, and it is trying to make more transactions happen and more projects successful. We then analyze the moderating effects on backers' utilities. Besides *Trial Stage*, *log(SampleNum)* has positive moderating effects on backers' utilities, and *ReportScore* also has a positive moderating effect on backers' utilities at *Report Stage*. This result indicates that the number of samples and the feedback report's score are effective positive signals of project quality. The insignificant effect of *log(SampleNum)* in *Trial Stage* may be due to backers' strong will waiting for feedback reports and less attention on other things.

5.4 Counterfactual Analysis

The result of our structural model indicates that the sampling campaign benefits projects by providing them more opportunities to be noticed by backers, but it also lowers backers' utility obtained from supporting the project. However, the integrated effect of the sampling campaign on the demand of projects is still unclear. In this subsection, we perform a counterfactual analysis by simulations to find the true impact of the sampling campaign on demand.

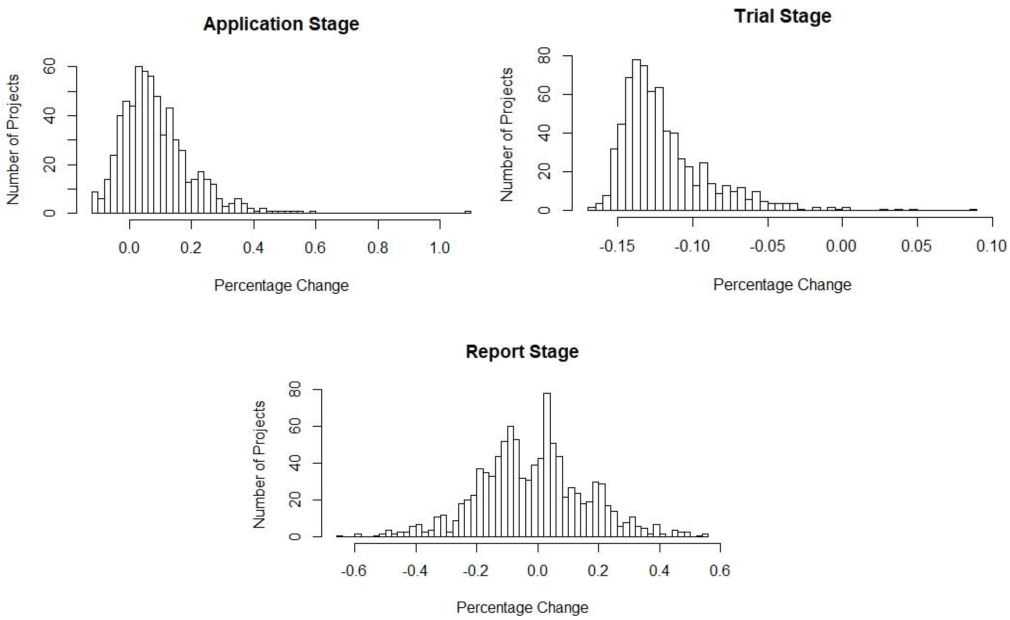

Figure 5.4. Histograms of percentage changes in three stages

The simulation is performed for every project joining the sampling campaign when the sampling campaign is ongoing. We use the result of our heterogeneous model to make our simulation more accurate. We simulate the case where the project does not join the sampling campaign by recalculating the probability of the project to be noticed φ_{pt}, regenerating consideration sets of every sampled backer, then recalculating the conditional market shares and the aggregated simulated market share. We compare the new market share where the project does not take part in the sampling campaign with the real market share to find the impact of the sampling campaign on demand. The result of the simulation is illustrated in Figure 5.4 and Table 5.3.

Our result shows that the demand for most projects is boosted in *Application Stage* and plunges in *Trial Stage*. In *Report Stage*, the mean effect is approximately zero. The effect in *Application Stage* is positive since that the sampling campaign catches the attention of many backers at this stage but does not harm their utilities. The negative effect in *Trial Stage* results from backers' strategic waiting for feedback reports. In *Report Stage*, the signaling effect of the sampling campaign still works to largely neutralize the negative effect of vanishing strategic delays of backers.

Table 5.3. Simulated Impact of Sampling Campaign on Demand

Sampling Stage	Mean	Min	25%	Median	75%	Max
Application	8.78%	-11.90%	0.92%	6.66%	14.31%	108.59%
Trial	-11.79%	-16.96%	-13.92%	-12.64%	-10.65%	8.65%
Report	-1.87%	-64.40%	-13.35%	-1.82%	8.38%	54.88%

Our counterfactual analysis can also provide critical managerial implications to founders and platforms. As sampling projects benefit from raising the awareness of backers and are harmed by backers' lowered utility towards supporting them, the relative scale of increased backers noticing the projects and decreased utility is important. In our data, raising the awareness of more backers dominates slightly lowered utility in *Application Stage*, and the opposite is true in *Trial Stage*. This is not always the case on other platforms. Fundraisers and platforms should strive to increase the number of backers aware of the project to fight against the negative effect on backers' utility.

5.5 Conclusions

The online crowdfunding market has been developing fast in recent years. However, the market is with relatively high uncertainty, and the problem of asymmetric information needs to be addressed. Backers resort to observational learning and follow the choice of the herd to reduce uncertainty, while crowdfunding platforms design various policies to reveal more information about projects to backers. Among these policies, free product sampling is a unique one that has not yet been studied and may become a double-edged sword. It offers fundraisers the opportunity to draw the attention of more backers and signal the high quality of their projects, but it also provides backers with feedback reports from peer backers, which serves as valuable information about the project and may delay their decision-making process. In this paper, we investigate the various mechanisms behind sampling campaigns on crowdfunding platforms and the impact of product sampling on the demand of projects.

Using a rich dataset from JD crowdfunding, we build a structural model considering both sides of the market. We relax the full-attention assumption of backers leveraging the consideration set model. We address the issue of selection bias by correlating the demand side and supply side of the market. We find that the sampling campaign is beneficial for projects to raise the awareness of backers on them in all three stages but lowers backers' utility of supporting the project in the latter two stages, as backers are strategically waiting for feedback reports in *Trial Stage* and find themselves overestimating the quality of the project in *Report Stage*. We also find the empirical evidence of two quality signals, number of samples and feedback reports' score, in eligible stages. Our counterfactual analysis shows that the integrated effect of the sampling campaign on demand is positive in *Application Stage* and negative in *Trial Stage* for most projects. The impact of product sampling on demand in *Report Stage* varies substantially across projects, and the number of samples and the feedback report's score play important roles in causing the variation. Fundraisers and the platforms should work together to make more backers aware of the sampling campaigns and hence projects joining the sampling campaigns to boost the demand. The platform may consider strategies to alleviate backers' strategic delay of investment when they are waiting for evaluations from peer backers through feedback reports.

Our research contributes to the literature in several aspects. First, our study enriches the body of studies on crowdfunding. Specifically, we study the impact of the sampling campaign as a tool for platforms to alleviate the severe information asymmetry problem in the crowdfunding market. We delve into the mechanisms behind this promotional strategy on crowdfunding platforms and disentangle its effect on backers' awareness from its impact on backers' utilities. We combine the phenomena discovered in various fields, including the signaling effect (Chevalier and Mayzlin 2006; Duan et al. 2008; Liu et al. 2021), customers’ awareness (Goeree 2008), and consumers’ strategic delay (Greenleaf and Lehmann 1995; Levin et al. 2009; Yu et al. 2016), to explain our findings. Second, we contribute to the literature on product sampling. We document the performance of product sampling, an effective promotional strategy, in the context of the crowdfunding market. We find that product sampling is not working well as expected due to the scarcity of evaluations from peer backers, which indicates that the effectiveness of product sampling should be examined in different contexts as mechanisms may be different. Third, we identify two quality signals in the crowdfunding market: the number of samples and the feedback report’s score. This adds to the literature of signaling effect in the online context (Chevalier and Mayzlin 2006; Duan et al. 2008; Ho et al. 2017). Finally, we make methodological contributions to IS literature by using a consideration set model to account for limited attention from the demand side of the market. Our model can be adapted in the research of online platforms where consumers are impossible to notice all eligible options.

Our study can potentially be extended in several ways in future research. First, we cannot observe the true consideration set of every backer. A direction of extension on our model is to use observations on backers’ or consumers’ browsing history to characterize their decision processes more accurately. Second, if the investment history of backers is available, future studies can focus on the heterogeneous effect of sampling campaigns on different backers. We hope that our study sheds light on the mechanisms of product sampling in crowdfunding markets and provides insights for future studies investigating related contexts.

Chapter 6. CONCLUDING REMARKS

In recent years, online platforms explored various promotional strategies to improve their performances. In my dissertation, I focus on several promotional strategies on different online platforms to examine their impacts. My first essay identifies the signaling effect of sampling size in online physical goods sampling on e-commerce platforms. My second essay studies and compares different inventory information disclosure policies in the context of online sales with limited inventory on e-commerce platforms. My third work investigates the effect of sampling strategies on crowdfunding projects for fundraisers on crowdfunding platforms. My three works delve deep into the mechanisms of promotional strategies to find the true influence of the strategies.

My first essay focuses on online physical goods sampling on e-commerce platforms. Building a structural model characterizing the behavior of sellers and buyers, my work provides empirical evidence that sampling size serves as a positive quality signal of the product. Moreover, I find that product types (search, experience, and credence) act as moderators of the signaling effect of sampling size. I further perform a policy simulation on the sampling threshold, and I show that increasing the threshold would decrease average sales and number of e-tailers willing to join the sampling campaign. Platforms can leverage the results in this essay to adjust the size of their sampling campaigns.

My second essay studies inventory disclosure policies in online sales of limited inventory. My work investigates the best timing of disclosing the inventory information to maximize the expected sales. I consider four policies: “always disclose”, “never disclose”, the fixed threshold policy, and the time-dependent threshold policy. I proved that under mild conditions which are consistent with the empirical findings, time-dependent threshold policy is the optimal policy. I use dynamic programming to develop efficient algorithms to optimize policy parameters for two threshold policies. Through a numerical study, I explore the relative performances of different policies when the importance of herding effect and scarcity effect varies. My work provides not only effective and efficient algorithms for platforms to implement on their flash sales, but advice on selecting inventory disclosure policies as well.

My third essay also considers sampling campaigns, but in the context of crowdfunding platforms. Combining the consideration set model and the aggregated demand model, I capture two underlying channels through which sampling affects the focal project. I find that the sampling campaign can raise the awareness of more backers on the focal project, but it also lowers backers' utility of supporting the project in latter stages of the campaign as backers' strategic delay waiting for valuable feedback reports and reports not meeting the expectations of backers come into play. I also find that the score of feedback report and number of free samples act as valid quality signals of the project. Finally, the counterfactual analysis shows that the demand rises in the first stage, drops in the second stage, and does not change much in the last stage of the sampling campaign. Crowdfunding platforms and fundraisers can learn from my work that sampling is not always good for projects, and they should probably think of better ways to reveal project information.

The contribution of my dissertation to the literature is two-fold. First, I empirically identify the effects of promotional strategies on online platforms and the underlying mechanisms in promotional campaigns. Second, I construct analytical models to find optimal strategies in promotional campaigns for platforms and sellers. In short, my essays investigate promotional strategies on online platforms both empirically and analytically.

BIBLIOGRAPHY

[1] Abrahams, A. S., W. Fan, G. A. Wang, Z. Zhang, J. Jiao, 2015. An integrated text analytic framework for product defect discovery. *Production and Operations Management*, 24(6): 975-990.

[2] Aflaki, A. and Swinney, R., 2019. Inventory integration with rational consumers. *Available at SSRN: https://ssrn.com/abstract=2842182.*

[3] Agrawal, A., Catalini, C., and Goldfarb, A., 2014. Some Simple Economics of Crowdfunding. *Innovation Policy and the Economy*. 14: 63-97.

[4] Akerlof, G. A., 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*. 84(3): 488-500.

[5] Alexa. Top Sites in China. Retreived April 1, 2020, https://www.alexa.com/topsites/countries/CN.

[6] Aviv, Y., Wei, M. M., and Zhang, F., 2019. Responsive pricing of fashion products: The effects of demand learning and strategic consumer behavior. *Management Science*, 65(7): 2982-3000.

[7] Ba, S., Y. Jin, X. Li, X. Lu., 2020. One Size Fits All? The Differential Impact of Online Reviews and Coupons. *Production and Operations Management*, 29(10): 2403-2424.

[8] Balakrishnan, A., S. Sundaresan, B. Zhang, 2014. Browse-and-Switch: Retail-Online Competition under Value Uncertainty. *Production and Operations Management*, 23(7): 1129-1145.

[9] Bawa, K., and Shoemaker, R., 2004. The effects of free sample promotions on incremental brand sales. *Marketing Science*. 23(3): 345-363.

[10] Bell, D., S. Gallino, S., A. Moreno, 2015. Showrooms and Information Provision in Omni-channel Retail. *Production and Operations Management*, 24(3): 360-362.

[11] Belleflamme, P., Lambert, T., and Schwienbacher, A., 2014. Crowdfunding: Tapping the right crowd. *Journal of business venturing*. 29(5): 585-609.

[12] Berry, S., J. Levinsohn, A. Pakes, 1995. Automobile prices in market equilibrium. *Econometrica*, 63(4): 841-890.

[13] Bøg, M., Harmgart, H., Huck, S., and Jeffers, A. M., 2012. Fundraising on the Internet. *Kyklos*, 65(1): 18-30.

[14] Burtch, G., Ghose, A., and Wattal, S., 2013. An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information Systems Research*. 24(3): 499-519.

[15] Burtch, G., Ghose, A., and Wattal, S., 2016. Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research*. 27(3): 478-496.

[16] Burtch, G., Hong, Y., and Liu, D., 2018. The role of provision points in online crowdfunding. *Journal of Management Information Systems*. 35(1): 117-144.

[17] Cachon, G. P. and Swinney, R., 2011. The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. *Management science*, 57(4): 778-795.

[18] Calvo, E., Cui, R., and Wagner, L., 2020. Disclosing product availability in online retail. *Manufacturing & Service Operations Management*, forthcoming.

[19] Cao, Z., K.-L. Hui, H. Xu., 2018. When discounts hurt sales: The case of daily-deal markets. *Information Systems Research*, 29(3): 567-591.

[20] Cascino, S., Correia, M., and Tamayo, A., 2019. Does consumer protection enhance disclosure credibility in reward crowdfunding? *Journal of Accounting Research*. 57(5): 1247-1302.

[21] Chakraborty, S., and Swinney, R., 2021. Signaling to the Crowd: Private Quality Information and Rewards-Based Crowdfunding. *Manufacturing & Service Operations Management*. 23(1): 155-169.

[22] Chellappa, R. K., S. Shivendu, 2005. Managing piracy: Pricing and sampling strategies for digital experience goods in vertically segmented markets. *Information Systems Research*, 16(4): 400-417.

[23] Chen, S., Lei, J., and Moinzadeh, K., 2021. When to lock the volatile input price? Procurement of commodity components under different pricing schemes. *Manufacturing & Service Operations Management*, forthcoming.

[24] Chen, Y., J. Xie, 2008. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3): 477-491.

[25] Cheng, H. K., S. Li, Y. Liu, 2015. Optimal software free trial strategy: Limited version, time-locked, or hybrid? *Production and Operations Management*, 24(3): 504-517.

[26] Cheng, H. K., and Liu, Y., 2012. Optimal software free trial strategy: The impact of network externalities and consumer uncertainty. *Information Systems Research*. 23(2): 488-504.

[27] Cheng, H. K., Q. C. Tang, 2010. Free trial or no free trial: Optimal software product design with network effects. *European Journal of Operational Research*, 205(2): 437-447.

[28] Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3): 345-354.

[29] Corazzini, L. and Greiner, B., 2007. Herding, social preferences and (non-) conformity. *Economics Letters*, 97(1): 74-80.

[30] Courtney, C., Dutta, S., and Li, Y., 2017. Resolving Information Asymmetry: Signaling, Endorsement, and Crowdfunding Success. *Entrepreneurship Theory and Practice*. 41(2): 265-290.

[31] Cui, R., Zhang, D. J., and Bassamboo, A., 2019. Learning from inventory availability information: Evidence from field experiments on Amazon. *Management Science*, 65(3): 1216-1235.

[32] Darby, M. R., E. Karni, 1973. Free competition and the optimal amount of fraud. *The Journal of Law and Economics*, 16(1): 67-88.

[33] Dellarocas, C., 2003. Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Management Science*, 49(10): 1407-1424.

[34] Dellarocas, C., 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, 52(10): 1577-1593.

[35] Duan, W., B. Gu, A. B. Whinston, 2008. Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems*, 45(4): 1007-1016.

[36] Eisend, M., 2008. Explaining the impact of scarcity appeals in advertising: The mediating role of perceptions of susceptibility. *Journal of Advertising*, 37(3): 33-40.

[37] Erdem, T. and Keane, M.P., 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, 15(1): 1-20.

[38] Ferreira, K. J., Lee, B. H. A., and Simchi-Levi, D., 2016. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1): 69-88.

[39] Ferreira, K. J., Simchi-Levi, D., and Wang, H., 2018. Online network revenue management using Thompson sampling. *Operations research*, 66(6): 1586-1602.

[40] Gallino, S., A. Moreno, 2018. The value of fit information in online retail: Evidence from a randomized field experiment. *Manufacturing & Service Operations Management*, 20(4): 767-787.

[41] Gao, F., X. Su, 2017. Online and offline information for omnichannel retailing. *Manufacturing & Service Operations Management*, 19(1): 84-98.

[42] Ghose, A., P. G. Ipeirotis, B. Li, 2014. Examining the impact of ranking on consumer behavior and search engine revenue. *Management Science*, 60(7): 1632-1654.

[43] Goeree, M. S., (2008). Limited information and advertising in the US personal computer industry. *Econometrica*. 76(5): 1017-1074.

[44] Greenleaf, E. A., and Lehmann, D. R., 1995. Reasons for substantial delay in consumer decision making. *Journal of Consumer Research*. 22(2): 186-199.

[45] Heiman, A., McWilliams, B., Shen, Z., and Zilberman, D., 2001. Learning and forgetting: Modeling optimal product sampling over time. Management Science. 47(4): 532-546.

[46] Hemer, J., 2011. *A snapshot on crowdfunding*. Arbeitspapiere Unternehmen und Region.

[47] Ho, Y.-C., Wu, J., and Tan, Y., 2017. Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research*. 28(3): 626-642.

[48] Jain, D., V. Mahajan, E. Muller, 1995. An approach for determining optimal product sampling for the diffusion of a new product. *Journal of Product Innovation Management*, 12(2): 124-135.

[49] Jamieson, L. F., F. M. Bass, 1989. Adjusting stated intention measures to predict trial purchase of new products: A comparison of models and methods. *Journal of Marketing Research*, 26(3): 336-345.

[50] Jiang, Y., Ho, Y.-C., Yan, X., and Tan, Y., 2018. Investor platform choice: herding, platform attributes, and regulations. *Journal of Management Information Systems*. 35(1): 86-116.

[51] Jung, J.M. and Kellaris, J.J., 2004. Cross-national differences in proneness to scarcity effects: The moderating roles of familiarity, uncertainty avoidance, and need for cognitive closure. *Psychology & Marketing*, 21(9): 739-753.

[52] Kempf, D. S., R. E. Smith, 1998. Consumer processing of product trial and the influence of prior advertising: A structural modeling approach, *Journal of Marketing Research*, 35(3): 325-338.

[53] Khern-am-nuai, W., K. Kannan, H. Ghasemkhani, 2018. Extrinsic versus intrinsic rewards for contributing reviews in an online platform, *Information Systems Research*, 29(4): 871-892.

[54] Kihlstrom, R. E., M. H. Riordan, 1984. Advertising as a Signal. *Journal of Political Economy*, 92(3): 427-450.

[55] Kirmani, A., A. R. Rao, 2000. No pain, no gain: A critical review of the literature on signaling unobservable product quality. *Journal of Marketing*, 64(2): 66-79.

[56] Klein, L. R., 1998. Evaluating the potential of interactive media through a new lens: Search versus experience goods. *Journal of Business Research*, 41(3): 195-203.

[57] Koning, R., and Model, J., 2013. Experimental Study of Crowdfunding Cascades: When Nothing is Better than Something. *Available at SSRN: https://papers.ssrn.com/abstract=2308161.*

[58] Ku, H.H., Kuo, C.C. and Kuo, T.W., 2012. The effect of scarcity on the purchase intentions of prevention and promotion motivated consumers. *Psychology & Marketing*, 29(8): 541-548.

[59] Küçükgül, C., Özer, Ö. and Wang, S., 2021. Engineering social learning: Information design of time-locked sales campaigns for online platforms. *Available at SSRN: https://papers.ssrn.com/abstract=3493744.*

[60] Kumar, N., L. Qiu, S. Kumar, 2018. Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies. *Information Systems Research*, 29(4): 849-870.

[61] Kunz, M. M., Bretschneider, U., Erler, M., and Leimeister, J. M., 2017. An empirical investigation of signaling in reward-based crowdfunding. *Electronic Commerce Research*. 17(3): 425-461.

[62] Kuppuswamy, V., and Bayus, B. L., 2018. *Crowdfunding Creative Ideas: The Dynamics of Project Backers*. The Economics of Crowdfunding: Startups, Portals and Investor Behavior, eds. D. 151-182.

[63] Lee, S.-Y., L. Qiu, A. Whinston, 2018. Sentiment manipulation in online platforms: An analysis of movie tweets. *Production and Operations Management*, 27(3): 393-416.

[64] Lee, Y.-J., Y. Tan, 2013. Effects of different types of free trials and ratings in sampling of consumer software: An empirical study. *Journal of Management Information Systems*, 30(3): 213-246.

[65] Letizia, P., M. Pourakbar, T. Harrison, 2018. The Impact of Consumer Returns on the Multichannel Sales Strategies of Manufacturers. *Production and Operations Management*, 27(2): 323-349.

[66] Levin, Y., McGill, J., and Nediak, M., 2009. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management science*. 55(1): 32-46.

[67] Li, C. L. and Kouvelis, P., 1999. Flexible and risk-sharing supply contracts under price uncertainty. *Management Science*, 45(10): 1378-1398.

[68] Li, L., S. Tadelis, X. Zhou, 2020. Buying reputation as a signal of quality: Evidence from an online marketplace. *The RAND Journal of Economics*, 51(4): 965-988.

[69] Lin, Z., Y. Zhang, Y. Tan, 2019. An empirical study of free product sampling and rating bias. *Information Systems Research*, 30(1): 260-275.

[70] Lipsman, A. and Liu, C., 2020. US Ecommerce 2020: Coronavirus boosts e-commerce forecast and will accelerate channel-shift. *eMarketer, https://www.emarketer.com/content/us-ecommerce-2020*

[71] Liu, C. Z., Y. A. Au, H. S. Choi, 2014. Effects of freemium strategy in the mobile app market: An empirical study of google play. *Journal of Management Information Systems,* 31(3): 326-354.

[72] Liu, Z., Lin, Z., Zhang, Y., and Tan, Y, 2021. The Signaling Effect of Sampling Size in Physical Goods Sampling Via Online Channels. *Production and Operations Management*, 31(2): 529-546.

[73] Luca, M., G. Zervas, 2016. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, 62(12): 3412-3427.

[74] Mavlanova, T., R. Benbunan-Fich, M. Koufaris. 2012, Signaling theory and information asymmetry in online commerce. *Information & Management*, 49(5): 240-247.

[75] Mayzlin, D., Y. Dover, J. Chevalier, 2014. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8): 2421-2455.

[76] Mehra, A., R. L. Saha, 2018. Utilizing public betas and free trials to launch a software product. *Production and Operations Management*, 27(11): 2025-2037.

[77] Meng, Z., Y. Wang, Z. Lin, Y. Tan, 2020. The influence of platform inspection disclosure on sharing economy: Evidence from unstructured data. *Working paper. Available at SSRN: https://ssrn.com/abstract=3034113.*

[78] Milgrom, P., J. Roberts, J., 1986. Price and advertising signals of product quality. *Journal of Political Economy*, 94(4): 796-821.

[79] Nelson, P., 1970. Information and consumer behavior. *Journal of Political Economy*, 78(2): 311-329.

[80] Nelson, P., 1974. Advertising as information. *Journal of Political Economy*, 82(4): 729-754.

[81] Özer, O., Phillips, R., 2012. *The Oxford handbook of pricing management*. Oxford University Press.

[82] Papanastasiou, Y., and Savva, N., 2017. Dynamic pricing in the presence of social learning and strategic consumers. *Management Science*, 63(4): 919-939.

[83] Reid, J., 2013. The new generation of product sampling. *Retrieved July 25, 2021, https://retailtouchpoints.com/features/executive-viewpoints/the-new-generation-of-product-sampling*

[84] Rothschild, M. L., W. C. Gaidis., 1981. Behavioral learning theory: Its relevance to marketing and promotions. *Journal of marketing*, 45(2): 70-78.

[85] Sampler, 2020. Sampler raises $4 million as demand spikes amid pandemic. *Retrieved July 25, 2021, https://sampler.io/press/sampler-raises-4-million-as-demand-spikes-amid-pandemic.*

[86] Sayedi, A., and Baghaie, M., 2017. Crowdfunding as a Marketing Tool. *Available at SSRN: https://papers.ssrn.com/abstract=2938183.*

[87] Schultz, D. E., W. A. Robinson, L. Petrison, 1998. *Sales promotion essentials: the 10 basic sales promotion techniques-and how to use them, 3rd ed*. McGraw-Hill Professional.

[88] Sodero, A.C. and Rabinovich, E., 2017. Demand and revenue management of deteriorating inventory on the Internet: an empirical study of flash sales markets. *Journal of Business Logistics*, 38(3): 170-183.

[89] Spence, M., 1973. Job Market Signaling. *The Quarterly Journal of Economics*, 87(3): 355-374.

[90] Sun, H., L. Xu, 2018. Online reviews and collaborative service provision: A signal-jamming model. *Production and Operations Management*, 27(11): 1960-1977.

[91] Talluri, K.T., Van Ryzin, G. J., 2006. *The theory and practice of revenue management*. Springer Science & Business.

[92] Tan, X., Y. Wang, Y. Tan, 2019. Impact of live chat on purchase in electronic markets: The moderating role of information cues. *Information Systems Research*, 30(4): 1248-1271.

[93] Tan, Y., V. S. Mookerjee, 2005. Allocating spending between advertising and information technology in electronic retailing. *Management Science*, 51(8): 1236-1249.

[94] Taobao, About Taobao. *Retrieved April 1, 2020, https://www.taobao.com/about/intro.php*.

[95] Taylor, C., 2020. The time has come for e-commerce sampling. *Retreived July 25, 2021, https://www.forbes.com/sites/charlesrtaylor/2020/10/08/the-time-has-come-for-e-commerce-sampling/.*

[96] Wang, S., G. F. Özkan-Seely, 2018. Signaling product quality through a trial period. *Operations Research*, 66(2): 301-312.

[97] Wang, Y. and Tomlin, B., 2009. To wait or not to wait: Optimal ordering under lead time uncertainty and forecast updating. *Naval Research Logistics*, 56(8): 766-779.

[98] Wang, Y., P. Goes, Z. Wei, D. Zeng, 2019. Production of online word-of-mouth: Peer effects and the moderation of user characteristics. *Production and Operations Management*, 28(7): 1621-1640.

[99] Wei, X., Fan, M., You, W., and Tan, Y., 2021. An empirical study of the dynamic and differential effects of prefunding. *Production and Operations Management*, (30)5: 1331-1349.

[100] Wells, J. D., J. S. Valacich, T. J. Hess, 2011. What signal are you sending? How website quality influences perceptions of product quality and purchase intentions. *MIS quarterly*, 35(2): 373-396.

[101] Yang, L., Wang, Z. and Hahn, J., 2020. Scarcity strategy in crowdfunding: An empirical exploration of reward limits. *Information Systems Research*, 31(4): 1107-1131.

[102] Yang, M., Z. Zheng, V. Mookerjee. 2019. Prescribing response strategies to manage customer opinions: a stochastic differential equation approach. *Information Systems Research*, 30(2): 351-374.

[103] Yin, R., Aviv, Y., Pazgal, A., and Tang, C.S., 2009. Optimal markdown pricing: Implications of inventory display formats in the presence of strategic customers. *Management Science*, 55(8): 1391-1408.

[104] Yu, M., Debo, L., and Kapuscinski, R., 2016. Strategic waiting for consumer-generated quality information: Dynamic pricing of new experience goods. *Management Science*, 62(2): 410-435.

[105] Zhang, J., and Liu, P., 2012. Rational herding in microloan markets. *Management science*, (58)5: 892-912.

[106] Zhou, Z., Xiao, S., Ho, Y.-C. (Chad), and Tan, Y., 2021. Investor Learning in Crowdfunded Supply Chain Finance Markets. *Available at SSRN: https://papers.ssrn.com/abstract=3136240*.

APPENDIX A

A.1 DERIVING CONDITIONAL LIKELIHOOD

We derive the analytical form of conditional likelihood $P\left(\xi_j, \omega_j | u_j, d_j; \alpha, \beta, \sigma_\varepsilon, \mathbf{\Omega}\right)$ in this Appendix. We drop the dependency on u_j and d_j in the remaining part to simplify our expressions. Note that the distribution of $\left(\xi_j, \omega_j\right)$ is now dependent on $\left(u_j, d_j\right)$. Also, we denote $t_j = d_j - 1$ as the days that the sampling campaign goes without feedback reports.

To make the expressions clearer in the process of deriving likelihood, we simplify the expressions for the sales equation and for supply side. We modify the equation for sales as:

$$y_{jt} = x_{jt}^T \beta + \beta_q \ln\left(q_{jt} + 1\right) + \beta_r HasReport_{jt} + \beta_s AvgReportScore_{jt} + \beta_d DaysElapsed_{jt} + \xi_j + \varepsilon_{jt} \tag{A1.}$$

where $y_{jt} = ln\,(Sales_{jt} + 1)$ and x_{jt} are characteristics that can affect demand except $Size_{jt}$, $HasReport_{jt}$, $AvgReportScore_{jt}$, and $DaysElapsed_{jt}$. We further simplify $Size_j$ as q_j and $AvgReportScore_j$ as s_j. We modify the supply side equation as:

$$ln\left(\gamma_j\right) = z_j^T \alpha + \omega_j \tag{A2.}$$

where z_j are characteristics that can affect the trade-off factor γ_j. The joint distribution of ξ_j and ω_j is still denoted as a bivariate normal distribution conditional on u_j (with dependency on u_j dropped for simplicity):

$$\begin{pmatrix} \xi_j \\ \omega_j \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Omega}), \mathbf{\Omega} = \begin{pmatrix} \sigma_\xi^2 & \rho\sigma_\xi\sigma_\omega \\ \rho\sigma_\xi\sigma_\omega & \sigma_\omega^2 \end{pmatrix} \tag{A3.}$$

and the standard deviation of ε_{jt} is denoted as σ_ε.

We now turn to the calculation of the conditional likelihood function, which is:

$$P\left(\xi_j, \omega_j; \alpha, \beta, \sigma_\varepsilon, \mathbf{\Omega}\right) = \int_{\xi_j} L^D\left(\xi_j; \beta, \sigma_\varepsilon\right) \left(\int_{\underline{\omega}_j(\alpha,\beta,\sigma_\varepsilon)}^{\overline{\omega}_j(\alpha,\beta,\sigma_\varepsilon)} f_{\omega|\xi}\left(\omega_j | \xi_j; \mathbf{\Omega}\right) d\omega_j \right) f_\xi\left(\xi_j; \mathbf{\Omega}\right) d\xi_j \tag{A4.}$$

In the equation above, $L^D(\xi_j; \beta, \sigma_\varepsilon)$ is the likelihood of the sales when ξ_j is fixed. $f_{\omega|\xi}(\omega_j|\xi_j; \mathbf{\Omega})$ is the conditional probability density function of ω_j given ξ_j. $\underline{\omega}_j(\alpha, \beta, \sigma_\varepsilon)$ and $\overline{\omega}_j(\alpha, \beta, \sigma_\varepsilon)$ are lower and upper bounds of feasible ω_j. And $f_\xi(\xi_j; \mathbf{\Omega})$ is the probability density function of ξ_j.

We first calculate $\underline{\omega}_j(\alpha, \beta, \sigma_\varepsilon)$ and $\overline{\omega}_j(\alpha, \beta, \sigma_\varepsilon)$. These lower and upper bounds are derived from restrictions speculated from sampling sizes. To calculate restrictions, we need to calculate $\mathbb{E}(Sales_{jt})$. From our simplified expressions, we can find $\mathbb{E}(Sales_{jt})$ when the shop does not sample (Equation (A5)) and when the shop samples (Equation (A6)):

$$\mathbb{E}(Sales_{jt}|0) = \mathbb{E}\left(e^{ln\,(Sales_{jt}+1)}\right) - 1 = \mathbb{E}(e^{y_{jt}}) - 1 = e^{x_{jt}^T\beta + \xi_j + \frac{\sigma_\varepsilon^2}{2}} - 1 \tag{A5.}$$

$$\begin{aligned}\mathbb{E}(Sales_{jt}|q_j) &= \mathbb{E}\left(e^{ln(Sales_{jt}+1)}\right) - 1 = \mathbb{E}(e^{y_{jt}}) - 1 \\ &= e^{x_{jt}^T\beta + \beta_q\,ln(q_j+1) + \beta_d t + \mathbf{1}(t \geq d_j)\cdot(\beta_r + \beta_s s_j) + \xi_j + \frac{\sigma_\varepsilon^2}{2}} - 1\end{aligned} \tag{A6.}$$

Here we further simplify the depreciation factor e^{β_d} as δ. Next, we can find out lower and upper bounds for three types of shops (sampling and bounded, sampling and unbounded, and non-sampling).

For sampling and bounded shops, their restrictions are:

$$\sum_{t=1}^{T}\left(\mathbb{E}(Sales_{jt}|q_j+1) - \mathbb{E}(Sales_{jt}|q_j)\right) \leq r_j \tag{A7.}$$

$$\sum_{t=1}^{T}\left(\mathbb{E}(Sales_{jt}|q_j) - \mathbb{E}(Sales_{jt}|0)\right) \geq r_j \times q_j \tag{A8.}$$

Simplifying these inequalities, and using the characteristics of demand side at time $t = 1$ to approximate x_{jt}, we can derive following inequalities:

$$a_j + \xi_j \leq ln(r_j) \leq b_j + \xi_j \tag{A9.}$$

where a_j and b_j are:

$$a_j = S_0 + ln\left((q_j+2)^{\beta_q} - (q_j+1)^{\beta_q}\right) + ln\left(\frac{1-\delta^{t_j}}{1-\delta} + \frac{\delta^{t_j}(1-\delta^{T-t_j})}{1-\delta}\cdot e^{\beta_r+\beta_s s_j}\right) \tag{A10.}$$

$$b_j = S_0 + ln\left(\frac{1}{q_j}\left(\frac{1-\delta^{t_j}}{1-\delta}\cdot\left(\left(q_j+1\right)^{\beta_q}-1\right)+\frac{\delta^{t_j}\left(1-\delta^{T-t_j}\right)}{1-\delta}\cdot\left(\left(q_j+1\right)^{\beta_q}\cdot e^{\beta_r+\beta_s s_j}-1\right)\right)\right) \quad \text{(A11.)}$$

$$S_0 = x_{j1}^T\beta + \frac{\sigma_\varepsilon^2}{2} \quad \text{(A12.)}$$

Similarly, we can derive corresponding a_j and b_j for other two types of shops. For sampling and unbounded shops, we have:

$$a_j = S_0 + ln\left(\left(q_j+2\right)^{\beta_q}-\left(q_j+1\right)^{\beta_q}\right)+ln\left(\frac{1-\delta^{t_j}}{1-\delta}+\frac{\delta^{t_j}\left(1-\delta^{T-t_j}\right)}{1-\delta}\cdot e^{\beta_r+\beta_s s_j}\right) \quad \text{(A13.)}$$

$$b_j = S_0 + ln\left(\left(q_j+1\right)^{\beta_q}-{q_j}^{\beta_q}\right)+ln\left(\frac{1-\delta^{t_j}}{1-\delta}+\frac{\delta^{t_j}\left(1-\delta^{T-t_j}\right)}{1-\delta}\cdot e^{\beta_r+\beta_s s_j}\right) \quad \text{(A14.)}$$

and for non-sampling shops, we have:

$$a_j = S_0 + ln\left(\frac{1}{q_j}\left(\frac{1-\delta^{t_j}}{1-\delta}\cdot\left(\left(q_j+1\right)^{\beta_q}-1\right)+\frac{\delta^{t_j}\left(1-\delta^{T-t_j}\right)}{1-\delta}\cdot\left(\left(q_j+1\right)^{\beta_q}\cdot e^{\beta_r+\beta_s s_j}-1\right)\right)\right) \quad \text{(A15.)}$$

$$b_j = +\infty \quad \text{(A16.)}$$

Now we have $a_j + \xi_j \le ln(r_j) \le b_j + \xi_j$ for all shops, and we can plug in the expression of $ln(r_j)$ to get restrictions on ω_j:

$$\underline{\omega}_j(\alpha,\beta,\sigma_\varepsilon) = a_j + \xi_j - z_j^T\alpha \quad \text{(A17.)}$$

$$\overline{\omega}_j(\alpha,\beta,\sigma_\varepsilon) = b_j + \xi_j - z_j^T\alpha \quad \text{(A18.)}$$

Next, we calculate $L_D(\xi_j;\alpha,\beta,\sigma_\varepsilon)$. Suppose that we can observe T_j days of sales data for shop, and define $\delta_{jt} = y_{jt} - x_{jt}^T\beta$, we can get:

$$L_D(\xi_j;\alpha,\beta,\sigma_\varepsilon) = \prod_{t=1}^{T_j}\frac{1}{\sigma_\varepsilon}\varphi\left(\frac{y_{jt}-x_{jt}^T\beta-\xi_j}{\sigma_\varepsilon}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right)^{T_j} exp\left(-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T_j}\left(\delta_{jt}-\xi_j\right)^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right)^{T_j} exp\left(-\frac{T_j}{2\sigma_\varepsilon^2}\left(\left(\xi_j - \overline{\delta_{Jt}}\right)^2 + \overline{\left(\delta_{Jt}\right)^2} - \left(\overline{\delta_{Jt}}\right)^2\right)\right) \tag{A19.}$$

where $\varphi(\cdot)$ is the probability density function of standard normal distribution.

As we have restrictions on ω_j and expression of $L_D(\xi_j; \alpha, \beta, \sigma_\varepsilon)$ now, we can take the integration. However, it is difficult to integrate using ξ_j and ω_j, as $\underline{\omega}_j(\alpha, \beta, \sigma_\varepsilon)$ and $\overline{\omega}_j(\alpha, \beta, \sigma_\varepsilon)$ are functions of ξ_j. Instead of integrating directly, we integrate by substitution. We substitute ξ_j and ω_j as follow:

$$\begin{cases} u = \xi_j + \omega_j \\ v = \xi_j - \omega_j \end{cases} \tag{A20.}$$

As $a_j + \xi_j - z_j^T \alpha \leq \omega_j \leq b_j + \xi_j - z_j^T \alpha$, we can get that $z_j^T \alpha - b_j \leq v \leq z_j^T \alpha - a_j$ and there are no restrictions on u. In this case, restrictions of v is not a function of u and it is easier to integrate. We can get the variance-covariance matrix of u and v by $\mathbf{\Omega}$**:**

$$\mathbf{\Omega_{uv}} = \begin{pmatrix} \sigma_\xi^2 + \sigma_\omega^2 + 2\rho\sigma_\xi\sigma_\omega & \sigma_\xi^2 - \sigma_\omega^2 \\ \sigma_\xi^2 - \sigma_\omega^2 & \sigma_\xi^2 + \sigma_\omega^2 - 2\rho\sigma_\xi\sigma_\omega \end{pmatrix} \tag{A21.}$$

and we can also calculate the determinant of $\mathbf{\Omega_{uv}}$ as $|\mathbf{\Omega_{uv}}| = 4(1-\rho^2)\sigma_\xi^2\sigma_\omega^2$. Now we can calculate likelihood of a shop as:

$$L_j(\alpha, \beta, \sigma_\varepsilon, \mathbf{\Omega}) = \int_{-\infty}^{+\infty} \int_{z_j^T\alpha - b_j}^{z_j^T\alpha - a_j} \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right)^{T_j} exp\left(-\frac{T_j}{2\sigma_\varepsilon^2}\left(\left(\frac{u+v}{2} - \overline{\delta_{Jt}}\right)^2 + \overline{\left(\delta_{Jt}\right)^2}\right.\right.$$

$$\left.\left. - \left(\overline{\delta_{Jt}}\right)^2\right)\right) \frac{1}{4\pi\sigma_\xi\sigma_\omega\sqrt{1-\rho^2}} exp\left(-\frac{1}{2}(u,v)\mathbf{\Omega_{uv}^{-1}}\begin{pmatrix} u \\ v \end{pmatrix}\right) dv\, du$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right)^{T_j} exp\left(-\frac{T_j}{2\sigma_\varepsilon^2}\left(\overline{\left(\delta_{Jt}\right)^2} - \left(\overline{\delta_{Jt}}\right)^2\right)\right) \frac{1}{4\pi\sigma_\xi\sigma_\omega\sqrt{1-\rho^2}}$$

$$\int_{-\infty}^{+\infty} \int_{z_j^T\alpha - b_j}^{z_j^T\alpha - a_j} exp\left(-\frac{1}{2}(u,v)\mathbf{\Omega_{uv}^{-1}}\begin{pmatrix} u \\ v \end{pmatrix} - \frac{T_j}{2\sigma_\varepsilon^2}\left(\frac{u+v}{2} - \overline{\delta_{Jt}}\right)^2\right) dv\, du \tag{A22.}$$

To calculate the integration in above expression, we try to transform the integrated function into a probability density function of a bivariate normal distribution. Suppose that:

$$(u-u_0, v-v_0)\boldsymbol{\Sigma}_{\boldsymbol{uv}}^{-1}\begin{pmatrix} u-u_0 \\ v-v_0 \end{pmatrix} = (u,v)\boldsymbol{\Omega}_{\boldsymbol{uv}}^{-1}\begin{pmatrix} u \\ v \end{pmatrix} + \frac{T_j}{\sigma_\varepsilon^2}\left(\frac{u+v}{2} - \overline{\delta_{Jt}}\right)^2 + D \tag{A23.}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{uv}}^{-1} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Then we can solve (a, b, c, u_0, v_0, D). We first calculate $\boldsymbol{\Omega}_{\boldsymbol{uv}}^{-1}$ as:

$$\boldsymbol{\Omega}_{\boldsymbol{uv}}^{-1} = \frac{1}{4(1-\rho^2)}\begin{pmatrix} \frac{1}{\sigma_\xi^2} + \frac{1}{\sigma_\omega^2} - \frac{2\rho}{\sigma_\xi \sigma_\omega} & \frac{1}{\sigma_\xi^2} - \frac{1}{\sigma_\omega^2} \\ \frac{1}{\sigma_\xi^2} - \frac{1}{\sigma_\omega^2} & \frac{1}{\sigma_\xi^2} + \frac{1}{\sigma_\omega^2} + \frac{2\rho}{\sigma_\xi \sigma_\omega} \end{pmatrix} \tag{A24.}$$

and we define:

$$\begin{cases} A = \frac{1}{\sigma_\xi^2} + \frac{1}{\sigma_\omega^2} - \frac{2\rho}{\sigma_\xi \sigma_\omega} \\ B = \frac{1}{\sigma_\xi^2} - \frac{1}{\sigma_\omega^2} \\ C = \frac{1}{\sigma_\xi^2} + \frac{1}{\sigma_\omega^2} + \frac{2\rho}{\sigma_\xi \sigma_\omega} \end{cases} \tag{A25.}$$

Then we can solve (a, b, c, u_0, v_0, D) as:

$$\begin{cases} a = \frac{1}{4}\left(\frac{T_j}{\sigma_\varepsilon^2} + \frac{A}{1-\rho^2}\right) \\ b = \frac{1}{4}\left(\frac{T_j}{\sigma_\varepsilon^2} + \frac{B}{1-\rho^2}\right) \\ c = \frac{1}{4}\left(\frac{T_j}{\sigma_\varepsilon^2} + \frac{C}{1-\rho^2}\right) \\ u_0 = \frac{\sigma_\xi^2 + \rho\sigma_\xi\sigma_\omega}{\sigma_\xi^2 + \frac{\sigma_\varepsilon^2}{T_j}}\overline{\delta_{Jt}} \\ v_0 = \frac{\sigma_\xi^2 - \rho\sigma_\xi\sigma_\omega}{\sigma_\xi^2 + \frac{\sigma_\varepsilon^2}{T_j}}\overline{\delta_{Jt}} \\ D = \frac{1}{\sigma_\xi^2 + \frac{\sigma_\varepsilon^2}{T_j}}\left(\overline{\delta_{Jt}}\right)^2 \end{cases} \tag{A26.}$$

and we can find $\boldsymbol{\Sigma}_{\boldsymbol{uv}}$ and its determinant:

$$\boldsymbol{\Sigma_{uv}} = \frac{(1-\rho^2)\sigma_\xi^2\sigma_\omega^2}{1+\frac{T_j\sigma_\xi^2}{\sigma_\varepsilon^2}}\begin{pmatrix} \frac{T_j}{\sigma_\varepsilon^2}+\frac{C}{1-\rho^2} & -\left(\frac{T_j}{\sigma_\varepsilon^2}+\frac{B}{1-\rho^2}\right) \\ -\left(\frac{T_j}{\sigma_\varepsilon^2}+\frac{B}{1-\rho^2}\right) & \frac{T_j}{\sigma_\varepsilon^2}+\frac{A}{1-\rho^2} \end{pmatrix} \tag{A27.}$$

$$|\boldsymbol{\Sigma_{uv}}| = \frac{4(1-\rho^2)\sigma_\xi^2\sigma_\omega^2}{1+\frac{T_j\sigma_\xi^2}{\sigma_\varepsilon^2}} \tag{A28.}$$

Now we can calculate likelihood of a shop as:

$$L_j(\alpha,\beta,\sigma_\varepsilon,\boldsymbol{\Omega}) = \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right)^{T_j}\sqrt{\frac{1}{1+\frac{T_j\sigma_\xi^2}{\sigma_\varepsilon^2}}}\,exp\left(-\frac{T_j}{2\sigma_\varepsilon^2}\left(\overline{\left(\delta_{jt}\right)^2}-\frac{\sigma_\xi^2}{\sigma_\xi^2+\frac{\sigma_\varepsilon^2}{T_j}}\left(\overline{\delta_{jt}}\right)^2\right)\right)$$

$$\frac{1}{2\pi|\boldsymbol{\Sigma_{uv}}|^{\frac{1}{2}}}\int_{-\infty}^{+\infty}\int_{z_j^T\alpha-b_j}^{z_j^T\alpha-a_j} exp\left(-\frac{1}{2}(u-u_0,v-v_0)\boldsymbol{\Sigma_{uv}^{-1}}\binom{u-u_0}{v-v_0}\right)dv\,du \tag{A29.}$$

Note that now the integration becomes a probability density function of a bivariate normal distribution. To calculate this integration, suppose that u' and v' follow this distribution, we can first calculate variance of v' from $\boldsymbol{\Sigma_{uv}}$ as:

$$\sigma_{v'}^2 = \frac{(1-\rho^2)\sigma_\xi^2\sigma_\omega^2}{1+\frac{T_j\sigma_\xi^2}{\sigma_\varepsilon^2}}\left(\frac{T_j}{\sigma_\varepsilon^2}+\frac{A}{1-\rho^2}\right) \tag{A30.}$$

Then we can calculate the likelihood of a shop as:

$$L_j(\alpha,\beta,\sigma_\varepsilon,\boldsymbol{\Omega}) = \left(\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\right)^{T_j}\sqrt{\frac{1}{1+\frac{T_j\sigma_\xi^2}{\sigma_\varepsilon^2}}}\,exp\left(-\frac{T_j}{2\sigma_\varepsilon^2}\left(\overline{\left(\delta_{jt}\right)^2}-\frac{\sigma_\xi^2}{\sigma_\xi^2+\frac{\sigma_\varepsilon^2}{T_j}}\left(\overline{\delta_{jt}}\right)^2\right)\right)$$
$$\left(\Phi\left(\frac{z_j^T\alpha-a_j-v_0}{\sigma_{v'}}\right)-\Phi\left(\frac{z_j^T\alpha-b_j-v_0}{\sigma_{v'}}\right)\right) \tag{A31.}$$

where $\Phi(\cdot)$ is the cumulated distribution function of standard normal distribution.

Using the approach above, we can calculate the integration in likelihood function analytically and derive an analytical likelihood function without integration. We use this expression of likelihood function directly to calculate the full log-likelihood $LL(\alpha, \beta, \sigma_\varepsilon, \mathbf{\Omega})$.

A.2 Results of First Stage Estimation

Table A.1. First Stage Estimation Using Instrumental Variables

	IV Estimation
Dependent Variable: *ln(Price)*	
ln(PairedPrice)	0.913*** (0.005)
Delivery	-0.101*** (0.019)
PromiseNum	0.027*** (0.004)
ln(ReviewVolume)	-0.021*** (0.002)
ShopRating	0.538*** (0.071)
ln(ShopRatingVolume)	-0.006** (0.003)
CategoryDummies	√
DayTypeDummies	√
Constant	-2.203*** (0.348)
Number of observations	7,076
R^2	0.8935
F-statistic	4,231

*Note: Standard errors in parentheses; Significant levels: *** p <0.01, ** p <0.05, * p <0.1*

A.3 RESULTS OF ROBUSTNESS CHECKS

Table A.2 shows our robustness check on duration of sampling effect. Table A.3 presents robustness checks on adding monthly sales as control variables (column (1)), distinguishing treatment effects of sampling size when the sampling campaign is “Ongoing” and “Completed” (column (2)) and using relative time effect model (column (3)). $T_{11\text{-}20}$ is the dummy indicating that the sampling campaign has started 11-20 days, and $T_{21\text{-}30}$ is the dummy indicating that the sampling campaign has started 21-30 days. Our main results still hold in all cases. For brevity, we only report the coefficients of important variables and leave other variables as *Controls* in the tables.

Table A.2. Effect of Sampling Sizes with Different Durations of Sampling Effect

	(1) 20 days	(2) 40 days	(3) 60 days
Demand Equation:			
ln(Size)	0.172*** (0.022)	0.180*** (0.022)	0.181*** (0.022)
ln(Price)	-0.242*** (0.037)	-0.259*** (0.036)	-0.260*** (0.036)
*ln(Price) * Sample*	0.186*** (0.019)	0.202*** (0.019)	0.203*** (0.019)
HasReport	-2.152*** (0.509)	-1.861*** (0.489)	-1.717*** (0.482)
ReportScore	0.468*** (0.105)	0.413*** (0.101)	0.384*** (0.099)
DaysElapsed	-0.023*** (0.003)	-0.025*** (0.003)	-0.025*** (0.003)
Search	0.363** (0.160)	0.391** (0.163)	0.397** (0.164)
Credence	0.186 (0.184)	0.203 (0.188)	0.208 (0.189)
DayTypeDummies	√	√	√
Controls	√	√	√
Constant	-0.006 (0.772)	0.128 (1.850)	0.117 (2.068)
σ_ξ	0.883*** (0.049)	0.907*** (0.050)	0.908*** (0.050)
σ_ε	0.835*** (0.007)	0.835*** (0.007)	0.836*** (0.007)
Report Equation:			
ln(Price)	-0.046** (0.022)	-0.049** (0.020)	-0.050** (0.022)
Search	0.081 (0.061)	0.085 (0.060)	0.087 (0.062)
Credence	-0.045 (0.068)	-0.041 (0.067)	-0.042 (0.068)
Controls	√	√	√
Constant	5.158*** (1.494)	5.220*** (1.437)	5.187*** (1.481)
σ_u	0.217*** (0.014)	0.218*** (0.014)	0.218*** (0.014)
$\rho_{\xi u}$	-0.014 (0.106)	0.000 (0.014)	0.006 (0.108)
Supply Equation:			
Search	1.471*** (0.267)	1.467*** (0.266)	1.468*** (0.265)
Credence	0.646** (0.308)	0.642** (0.309)	0.644** (0.308)
Controls	√	√	√
Constant	1.025** (0.451)	1.924*** (0.459)	2.353*** (0.461)
ReportDeviation	0.882*** (0.177)	1.001*** (0.172)	1.158*** (0.174)
σ_ω	1.385*** (0.090)	1.414*** (0.091)	1.413*** (0.092)
$\rho_{\xi\omega}$	0.879*** (0.026)	0.888*** (0.023)	0.888*** (0.023)
$\rho_{\omega u}$	0.098 (0.124)	0.115** (0.054)	0.122 (0.125)
Number of observations	7,076	7,076	7,076
AIC	19,033.60	19,019.66	19,018.76

*Note: Standard errors in parentheses; Significant levels: *** p <0.01, ** p <0.05, * p <0.1*

Table A.3. Effect of Sampling Sizes with Other Robustness Checks

	(1) Adding monthly sales	(2) Distinguishing "Ongoing" and "Completed"	(3) Relative time effect model
Demand Equation:			
ln(Size)	0.179*** (0.022)	0.114*** (0.024)	0.163*** (0.024)
ln(MonthlySales)	0.028** (0.012)	/	/
ln(Size) * Completed	/	0.175*** (0.020)	/
ln(Size) * $T_{11\text{-}20}$	/	/	0.120*** (0.024)
ln(Size) * $T_{21\text{-}30}$	/	/	0.140*** (0.033)
ln(Price)	-0.241*** (0.036)	-0.264*** (0.034)	-0.265*** (0.035)
*ln(Price) * Sample*	0.191*** (0.019)	0.203*** (0.018)	0.202*** (0.018)
HasReport	-1.973*** (0.496)	-2.368*** (0.567)	-2.607*** (0.571)
ReportScore	0.436*** (0.102)	0.500*** (0.116)	0.533*** (0.116)
DaysElapsed	-0.024*** (0.003)	-0.034*** (0.003)	-0.030*** (0.004)
Search	0.365** (0.159)	0.439*** (0.164)	0.426*** (0.164)
Credence	0.196 (0.183)	0.226 (0.190)	0.218 (0.189)
DayTypeDummies	√	√	√
Controls	√	√	√
Constant	0.106 (1.900)	0.326 (1.872)	0.270 (1.883)
σ_ξ	0.880*** (0.049)	0.918*** (0.049)	0.913*** (0.049)
σ_ε	0.836*** (0.007)	0.831*** (0.007)	0.835*** (0.007)
Report Equation:			
ln(Price)	-0.047** (0.022)	-0.051** (0.022)	-0.053** (0.022)
Search	0.084 (0.061)	0.090 (0.062)	0.092 (0.062)
Credence	-0.042 (0.068)	-0.039 (0.068)	-0.037 (0.068)
Controls	√	√	√
Constant	5.262*** (1.504)	5.199*** (1.497)	5.165*** (1.504)
σ_u	0.218*** (0.014)	0.218*** (0.014)	0.218*** (0.014)
$\rho_{\xi u}$	-0.002 (0.103)	0.010 (0.108)	0.012 (0.106)
Supply Equation:			
Search	1.444*** (0.268)	1.382*** (0.248)	1.365*** (0.227)
Credence	0.624** (0.310)	0.595** (0.288)	0.575** (0.287)
Controls	√	√	√
Constant	1.639*** (0.457)	2.368*** (0.465)	2.341*** (0.476)
ReportDeviation	0.917*** (0.175)	1.436*** (0.264)	1.424*** (0.269)
σ_ω	1.414*** (0.092)	1.319*** (0.080)	1.313*** (0.081)
$\rho_{\xi\omega}$	0.889*** (0.023)	0.915*** (0.019)	0.911*** (0.020)
$\rho_{\omega u}$	0.110 (0.120)	0.131 (0.124)	0.144 (0.121)
Number of observations	7,076	7,076	7,076
AIC	19,018.79	18945.87	18,999.82

*Note: Standard errors in parentheses; Significant levels: *** p <0.01, ** p <0.05, * p <0.1*

APPENDIX B

B.1 MATHEMATICAL PROOFS

Proof of Proposition 1. First, we show that the general form of solutions to Equations (1)-(2) is as follows:

$$Q_0(t|\zeta,j) = C_0 e^{-\int_\zeta^t \lambda_j(s)ds},$$
$$Q_i(t|\zeta,j) = e^{-\int_\zeta^t \lambda_{j+i}(s)ds}\left[\int_\zeta^t e^{\int_\zeta^s \lambda_{j+i}(x)dx}\,\lambda_{j+i-1}(s)Q_{i-1}(s|\zeta,j)ds\right], i = 1,\cdots,K-j,$$

where C_0 has a non-negative value such that the boundary conditions are satisfied. Note that,

$$\begin{aligned}
\frac{dQ_0(t|\zeta,j)}{dt} &= -\lambda_j(t)\cdot C_0 e^{-\int_\zeta^t \lambda_j(s)ds} = -\lambda_j(t)Q_0(t|\zeta,j) \\
\frac{dQ_i(t|\zeta,j)}{dt} &= -\lambda_{j+i}(t)Q_i(t|\zeta,j) + e^{-\int_\zeta^t \lambda_{j+i}(s)ds}\left[e^{\int_\zeta^t \lambda_{j+i}(x)dx}\lambda_{j+i-1}(t)Q_{i-1}(t|\zeta,j)\right] \\
&= -\lambda_{j+i}(t)Q_i(t|\zeta,j) + \lambda_{j+i-1}(t)Q_{i-1}(t|\zeta,j).
\end{aligned}$$

Furthermore, in order to satisfy the boundary conditions, we should have $C_0 = 1$ such that $Q_0(\zeta|\zeta,j) = 1$ and $Q_i(\zeta|\zeta,j) = 0$ for any $i \in \{1,\cdots,K-j\}$.

■

Proof of Proposition 2. In order to prove Equation (8) of this proposition, we use induction, substituting $\lambda_i(t) = a(i)\lambda_0(t)$ into Equation (5)-(6) of Proposition 1. Note that

$$Q_0(t|\zeta,j) = e^{-\int_\zeta^t a(j)\lambda_0(s)ds} = e^{-a(j)\Lambda(\zeta,t)}.$$

For the first step of the backward induction, we substitute $Q_0(t|\zeta,j)$ into the expression of $Q_1(t|\zeta,j)$. That is,

$$
\begin{aligned}
Q_1(t|\zeta,j) &= e^{-\int_\zeta^t \lambda_{j+1}(s)ds}\left[\int_\zeta^t e^{\int_\zeta^s \lambda_{j+1}(x)dx}\,\lambda_j(s)Q_0(s|\zeta,j)ds\right] \\
&= e^{-\int_\zeta^t a(j+1)\lambda_0(s)ds}\left[\int_\zeta^t e^{\int_\zeta^s a(j+1)\lambda_0(x)dx}\,a(j)\lambda_0(s)e^{-\int_\zeta^s a(j)\lambda_0(x)dx}ds\right] \\
&= e^{-\int_\zeta^t a(j+1)\lambda_0(s)ds}\left[\int_\zeta^t e^{\int_\zeta^s [a(j+1)-a(j)]\lambda_0(x)dx}\,a(j)\lambda_0(s)ds\right] \\
&= e^{-\int_\zeta^t a(j+1)\lambda_0(s)ds}\frac{a(j)}{a(j+1)-a(j)}\left[e^{\int_\zeta^t [a(j+1)-a(j)]\lambda_0(x)dx} - e^{\int_\zeta^\zeta [a(j+1)-a(j)]\lambda_0(x)dx}\right] \\
&= \frac{a(j)}{a(j+1)-a(j)}\left[e^{-\int_\zeta^t a(j)\lambda_0(x)dx} - e^{-\int_\zeta^t a(j+1)\lambda_0(x)dx}\right] \\
&= \frac{a(j)}{a(j+1)-a(j)}\left[e^{-a(j)\Lambda(\zeta,t)} - e^{-a(j+1)\Lambda_0(\zeta,t)}\right] \\
&= a(j)\left[\frac{e^{-a(j)\Lambda(\zeta,t)}}{a(j+1)-a(j)} + \frac{e^{-a(j+1)\Lambda_0(\zeta,t)}}{a(j)-a(j+1)}\right],
\end{aligned}
$$

which is consistent with Equation (8) when $i = 1$.

For the second step of the backward induction, we have

$$
\begin{aligned}
&Q_2(t|\zeta,j) = e^{-\int_\zeta^t \lambda_{j+2}(s)ds}\left[\int_\zeta^t e^{\int_\zeta^s \lambda_{j+2}(x)dx}\,\lambda_{j+1}(s)Q_1(s|\zeta,j)ds\right] \\
=\;& e^{-a(j+2)\Lambda(\zeta,t)}\left[\int_\zeta^t e^{a(j+2)\Lambda(\zeta,s)}\,a(j+1)\lambda_0(s)\cdot\frac{a(j)}{a(j+1)-a(j)}\left[e^{-a(j)\Lambda(\zeta,s)} - e^{-a(j+1)\Lambda(\zeta,s)}\right]ds\right] \\
=\;& e^{-a(j+2)\Lambda(\zeta,t)}\left[\int_\zeta^t \frac{a(j+1)a(j)}{a(j+1)-a(j)}\left[e^{[a(j+2)-a(j)]\Lambda(\zeta,s)} - e^{[a(j+2)-a(j+1)]\Lambda(\zeta,s)}\right]\lambda_0(s)ds\right] \\
=\;& e^{-a(j+2)\Lambda(\zeta,t)}\frac{a(j+1)a(j)}{a(j+1)-a(j)}\left[\frac{e^{[a(j+2)-a(j)]\Lambda(\zeta,t)}-1}{a(j+2)-a(j)} - \frac{e^{[a(j+2)-a(j+1)]\Lambda(\zeta,t)}-1}{a(j+2)-a(j+1)}\right] \\
=\;& a(j+1)a(j)\left\{\frac{e^{-a(j)\Lambda(\zeta,t)} - e^{-a(j+2)\Lambda(\zeta,t)}}{[a(j+2)-a(j)][a(j+1)-a(j)]} - \frac{e^{-a(j+1)\Lambda(\zeta,t)} - e^{-a(j+2)\Lambda(\zeta,t)}}{[a(j+2)-a(j+1)][a(j+1)-a(j)]}\right\} \\
=\;& a(j+1)a(j)\{\frac{e^{-a(j)\Lambda(\zeta,t)}}{[a(j+2)-a(j)][a(j+1)-a(j)]} + \frac{e^{-a(j+1)\Lambda(\zeta,t)}}{[a(j+2)-a(j+1)][a(j)-a(j+1)]} + \cdots \\
&\qquad\cdots + \frac{e^{-a(j+2)\Lambda(\zeta,t)}}{[a(j+1)-a(j+2)][a(j)-a(j+2)]}\},
\end{aligned}
$$

which is consistent with Equation (8) when $i = 2$.

The above calculation of cases $i = 1$ and $i = 2$ can be generalized further for any $i = 3, \cdots, K-j$. Suppose that $Q_i(t|\zeta,j)$ is given by Equation (8), and consider $Q_{i+1}(t|\zeta,j)$. We have

$$
\begin{aligned}
Q_{i+1}(t|\zeta,j) &= e^{-\int_\zeta^t \lambda_{j+i+1}(s)ds}\left[\int_\zeta^t e^{\int_\zeta^s \lambda_{j+i+1}(x)dx}\,\lambda_{j+i}(s)Q_i(s|\zeta,j)ds\right] \\
&= e^{-a(j+i+1)\Lambda(\zeta,t)}\left\{\int_\zeta^t e^{a(j+i+1)\Lambda(\zeta,s)}\,a(j+i)\lambda_0(s)Q_i(s|\zeta,j)ds\right\} \\
&= e^{-a(j+i+1)\Lambda(\zeta,t)}\left\{\int_\zeta^t\left(\prod_{k=0}^{i} a\,(j+k)\right)\left(\sum_{k=0}^{i}\frac{e^{[a(j+i+1)-a(j+k)]\Lambda(\zeta,s)}}{\prod_{0\le l\le i,l\ne k}(\,a(j+l)-a(j+k))}\right)\lambda_0(s)ds\right\}
\end{aligned}
$$

$$
= \quad e^{-a(j+i+1)\Lambda(\zeta,t)}\left(\prod_{k=0}^{i} a\,(j+k)\right)\left(\sum_{k=0}^{i}\frac{e^{[a(j+i+1)-a(j+k)]\Lambda(\zeta,t)}-1}{\prod_{0\le l\le i+1,l\ne k}(\,a(j+l)-a(j+k))}\right)
$$

$$
= \quad \left(\prod_{k=0}^{i} a\,(j+k)\right)\left(\sum_{k=0}^{i+1}\frac{e^{-a(j+k)\Lambda(\zeta,t)}}{\prod_{0\le l\le i+1,l\ne k}(\,a(j+l)-a(j+k))}\right),
$$

which is consistent with the general form of Equation (8) when i is replaced by $i+1$. Thus, we have completed the induction and proven that Equation (8) holds for any $i=1,\cdots,K-j$.

Last, substituting Equation (8) into $\Gamma(T|\zeta,j)=\sum_{i=1}^{K-j}\{\,i\cdot Q_i(T|\zeta,j)\}$, we obtain

$$
\begin{aligned}
\Gamma(T|\zeta,j) &= \sum_{i=0}^{K-j}\left\{i\cdot\left(\prod_{k=0}^{i-1} a\,(j+k)\right)\cdot\left(\sum_{k=0}^{i}\frac{e^{-a(j+k)\Lambda(\zeta,t)}}{\prod_{0\le l\le i,l\ne k}(\,a(j+l)-a(j+k))}\right)\right\} \\
&= \sum_{k=0}^{K-j}\left\{\sum_{i=k}^{K-j}\left[i\cdot\left(\prod_{m=0}^{i-1} a\,(j+m)\right)\cdot\left(\frac{e^{-a(j+k)\Lambda(\zeta,t)}}{\prod_{0\le l\le i,l\ne k}(\,a(j+l)-a(j+k))}\right)\right]\right\} \\
&= \sum_{k=0}^{K-j}\left\{e^{-a(j+k)\Lambda(\zeta,t)}\left[\sum_{i=k}^{K-j} i\left(\prod_{m=0}^{i-1} a\,(j+m)\right)\left(\frac{1}{\prod_{0\le l\le i,l\ne k}(\,a(j+l)-a(j+k))}\right)\right]\right\} \\
&= \sum_{i=0}^{K-j}\left\{e^{-a(j+i)\Lambda(\zeta,t)}\left[\sum_{k=i}^{K-j} k\left(\prod_{m=0}^{k-1} a\,(j+m)\right)\left(\frac{1}{\prod_{0\le l\le k,l\ne i}(\,a(j+l)-a(j+i))}\right)\right]\right\}.
\end{aligned}
$$

■

Proof of Theorem 1. We first show that the following statement is true: If it is optimal to disclose inventory at time t when j sales have been made until t, then it is always better to reveal than conceal inventory at any time $t'<t$ when j sales have been made until t'.

Suppose that at time τ instead of t, where $t-\tau=\varDelta$ is an infinitesimal time interval, the j-th sale occurs and the platform has not broadcasted its inventory. If the platform decides to reveal its inventory immediately at τ, then its expected future sales would be

$$\big(1-\lambda_j(\tau)\varDelta\big)\varGamma(T|t,j)+\lambda_j(\tau)\varDelta[\varGamma(T|t,j+1)+1]+o(\varDelta). \tag{B.1}$$

In contrast, if the platform decides not to reveal its inventory immediately, then its expected future sales would be

$$(1-\theta^{BD}(\tau)\varDelta)\varGamma(T|t,j)+\theta^{BD}(\tau)\varDelta[\varGamma(T|t,j+1)+1]+o(\varDelta). \tag{B.2}$$

In Equation (B.1) or Equation (B.2), with probability $(1-\lambda_j(\tau)\varDelta)$ or $(1-\theta^{BD}(\tau)\varDelta)$, respectively, there will be no additional sale made in $[\tau,t]$. Then, at time t, the platform will start to broadcast its inventory as it is assumed to be optimal to broadcast inventory levels if there are j sales until t, and, as a result, the expected future sales thereafter will be $\varGamma(T|t,j)$. Otherwise, with probability $\lambda_j(\tau)\varDelta$ or $\theta^{BD}(\tau)\varDelta$, respectively, there will be $j+1$ sales until t. In that case, the platform will also start to broadcast its inventory at t, and the expected future sales thereafter will be $\varGamma(T|t,j+1)$. The reason is that, if the platform finds itself better off revealing than concealing its inventory at time t with j sales, it must also find itself better off revealing than concealing its inventory at the same time t with more sales (e.g., $j+1$ sales) made, as, according to Assumption 2, $\beta(j+1,t)>\beta(j,t)$ and thus $\lambda_{j+1}(t)>\lambda_j(t)$.

Next, note that a necessary condition for revealing inventory information at time t with j sales is: $\lambda_j(t)\geq\theta^{BD}(t)$, which implies $\beta(j,t)\geq 1$. Then, according to Assumption 2, we have $\beta(j,\tau)>\beta(j,t)\geq 1$ given that $\tau<t$, which implies $\lambda_j(\tau)>\theta^{BD}(\tau)$ as well. In addition, it is clear that $\varGamma(T|t,j+1)+1>\varGamma(T|t,j)$. That is, we have $\lambda_j(\tau)>\theta^{BD}(\tau)$ and $\varGamma(T|t,j+1)+1\geq\varGamma(T|t,j)$, which together establish the fact that Equation (B.1) is greater than Equation (B.2), so the platform must be better off revealing its inventory at τ if there are j sales made, given that it is also optimal to do so at time $t = \tau + \Delta$. Using backward induction, the conclusion holds for any time $t' < t$.

Furthermore, since the above statement is true, there must exist a threshold t_j^* such that the platform should broadcast inventory information if and only if the j sales are made before t_j^*. The reason is as follows. Suppose that it will be better to reveal than conceal inventory information at certain time t when there are j sales made, then the platform should also reveal its inventory at any time $t' \leq t$ when the j-th sale occurs. Let t_j^* be the largest t such that the platform prefers revealing to concealing its inventory, given that there are j sales made. As a result, if the j-th sale occurs before t_j^*, then it is optimal to broadcast inventory; otherwise, it is optimal not to broadcast. Such a policy is exactly what the proposed time-dependent threshold policy describes.

Last, the monotonicity of t_j^*'s is a result of Assumption 1, which indicates that the platform cannot “flip-flop" its inventory disclosure decisions. Specifically, suppose that there exists certain j such that $t_{j+1}^* < t_j^*$. As long as the j-th sale occurs before t_j^*, the platform should have already revealed its inventory according to the proposed time-dependent policy. After that, even though the $(j+1)$-th sale may occur after t_{j+1}^*, it does not make a difference as the platform will not reverse its previous disclosure decision due to Assumption 1. In that case, indeed we can set $t_{j+1}^* = t_j^*$ without affecting the outcome of implementing the proposed policy.

∎

Proof of Proposition 3. Recall that $V^*(T|t,j)$ is the expected “sales-to-go" given that the platform does not disclose its inventory at the current time t when there are j sales. However, the platform may decide to disclose its inventory in the future.

The platform may start to broadcast its inventory upon the $(j+1)$-th sale in the future. The probability is $q_1(y|t,j)$ given that the $(j+1)$-th sale occurs at $[y - dy, y]$. In such a case, the expected “sales-to-go" includes the one additional sale that will occur at $[y-dy,y]$ and the expected sales thereafter, $\Gamma(T|y,j+1)$. Note that $y\in[t,t_{j+1}^*]$ because, otherwise, the platform will not optimally broadcast its inventory at $[y-dy,y]$ upon the occurrence of the $(j+1)$-th sale.

Similarly, the platform may start to broadcast its inventory levels upon the $(j+i)$-th sale in the future, where $i\in\{2,\cdots,K-j\}$. The probability is $q_i(y|t,j)$ given that the $(j+i)$-th sale occurs at $[y-dy,y]$. The expected “sales-to-go" includes the i additional sales that will occur during $[t,y]$ and the expected sales thereafter, $\Gamma(T|y,j+i)$. It is clear that $y\le t_{j+i}^*$ because, otherwise, the platform will not optimally disclose its inventory at $[y-dy,y]$ upon the occurrence of the $(j+i)$-th sale. We also note that $y>t_{j+i-1}^*$. The reason is as follows: suppose $y\le t_{j+i-1}^*$, then the $(j+i-1)$-th sale must have occurred before t_{j+i-1}^*, implying that the platform must have already disclosed its inventory upon the occurrence of the $(j+i-1)$-th sale at the latest.

Last, the platform may conceal its inventory until the end of the sale. The third term of Equation (17) gives the expected “sales-to-go" given that the platform conceals its inventory until the end of the sale period. Specifically, with probability $r_n(T|t,j)$, there will be n additional sales while the platform keeps concealing the inventory levels during $[t,T]$. Note that, since $t_K^*=T$, n cannot exceed $K-j-1$ as, otherwise, the platform would have revealed its inventory during $[t,T]$.

■

B.2 Numerical Study Extensions

Figure B.1 depicts how the maximum weight associated with the quality signal in the customers’ updating of their perceived quality, ρ_{max}, affects the four measures of improvements. We fix other parameters as default ones and let ρ_{max} varies from 0.1 to 0.9. We find that R_1, R_2, and R_4 are increasing in ρ_{max}, while R_3 has a peak between $\rho_{max} = 0.6$ and 0.7. The increasing pattern of R_1 and R_2 is intuitive, as the increase of ρ_{max} endows the customers with stronger ability to learn the quality of the product, and observational learning effect thus benefits the customers more. With stronger observational learning effect, "always disclose" policy performs better. Also, as "never disclose" policy dominates "always disclose" policy under this particular group of settings, the relative improvement of adopting the fixed threshold policy compared to the better of the two simple policies (i.e., the "never disclose" policy) increases. Moreover, R_4 is exhibiting a similar pattern with R_2, as the magnitude of R_3 is much smaller than that of R_2.

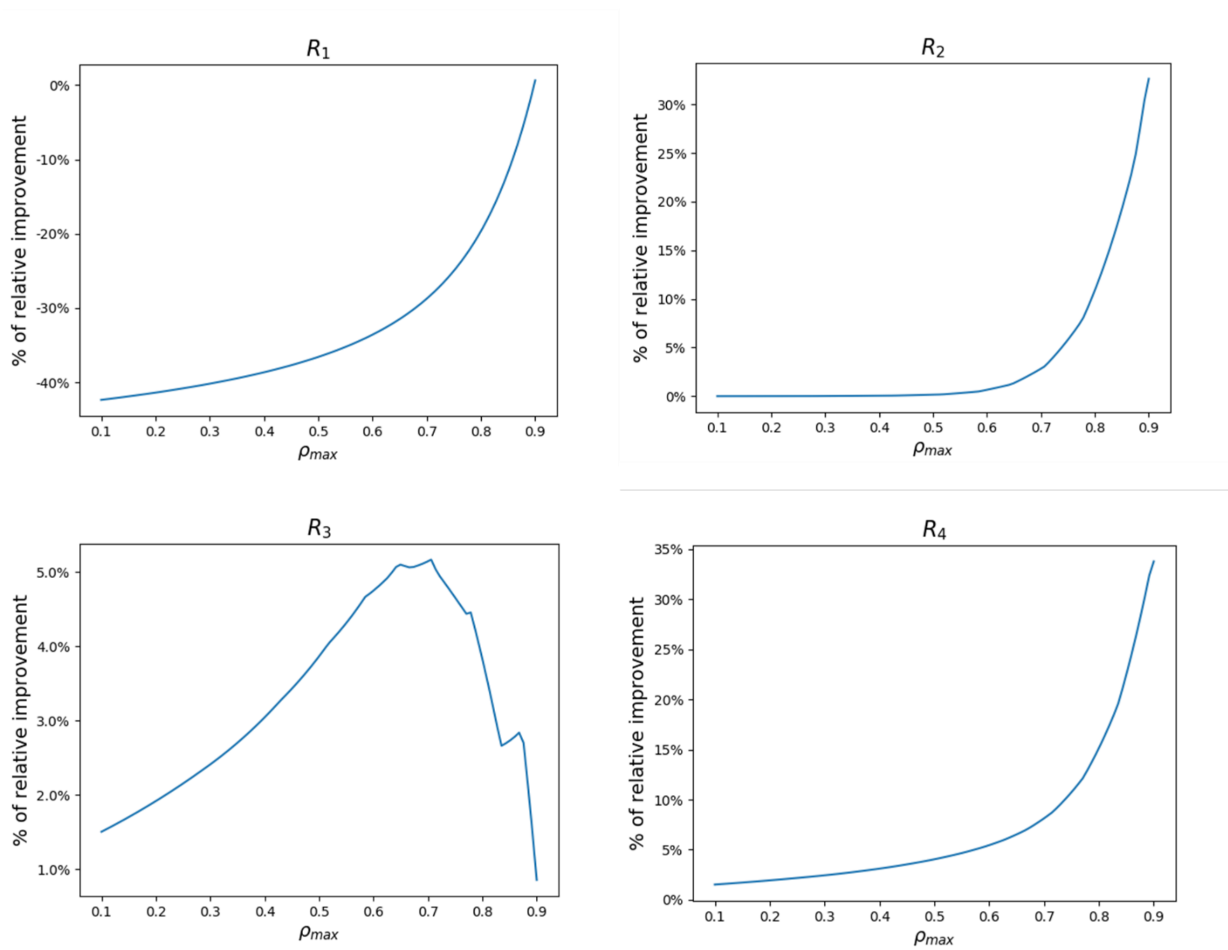

Figure B.1. Measures of improvements as ρ_{max} varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

To investigate the behavior of R_3, we again resort to the optimal fixed threshold and time-dependent thresholds. Figure B.2 illustrates the impact of ρ_{max} on the thresholds. As ρ_{max} increases, observational learning effect becomes more prominent. As a result, the optimal fixed threshold decreases and the optimal time thresholds associated with different levels of inventory sold increase, both encouraging the early disclosure of inventory information. As what we have discussed in the paper, the peak of R_3 is located at where the difference between the two threshold policies is the largest, which is between $\rho_{max} = 0.6$ and 0.7. At $\rho_{max} = 0.6$, for example, the optimal fixed threshold policy indicates that the platform should disclose inventory once 50% of its inventory is sold. In contrast, the optimal time-dependent threshold indicates that $t_j^* < 1$ for $j/K < 70\%$ and $t_j^* = 1$ for $j/K \geq 70\%$, which implies that the platform may disclose its inventory at any time before 70% of the inventory is sold and will disclose inventory after 70% of inventory is sold.

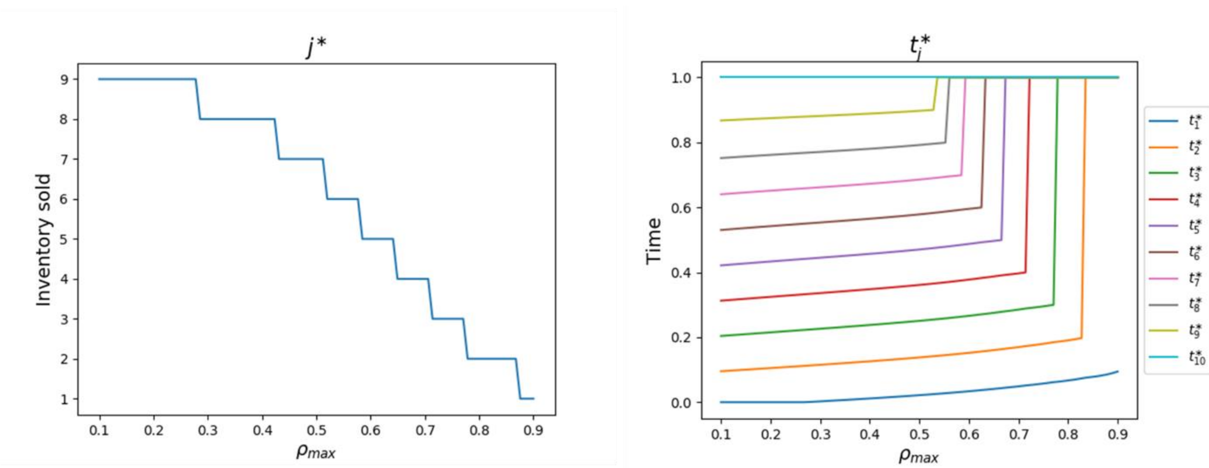

Figure B.2. The optimal thresholds of the fixed (left) and time-dependent (right) threshold policies as ρ_{max} varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

Figures B.3 and B.4 illustrate the impacts of q_{min} on the four measures of improvements and the associated thresholds. Other parameters are fixed at default values. We find that R_1 is increasing in q_{min} while R_2, R_3, and R_4 are decreasing in q_{min}. Increasing q_{min} reduces the heterogeneity of the customers' prior beliefs of product quality and undermines the importance of observational learning. Hence, the fixed and time-dependent threshold policies can improve less when q_{min} increases, and the optimal fixed threshold increases and the optimal time thresholds associated with different levels of inventory sold decrease with q_{min}, thus discouraging the platform to disclose inventory information early. Also, the increase of q_{min} raises customers' mean prior beliefs. As a result, the "always disclose" policy can attract more customers and close the gap between it and the "never disclose" policy.

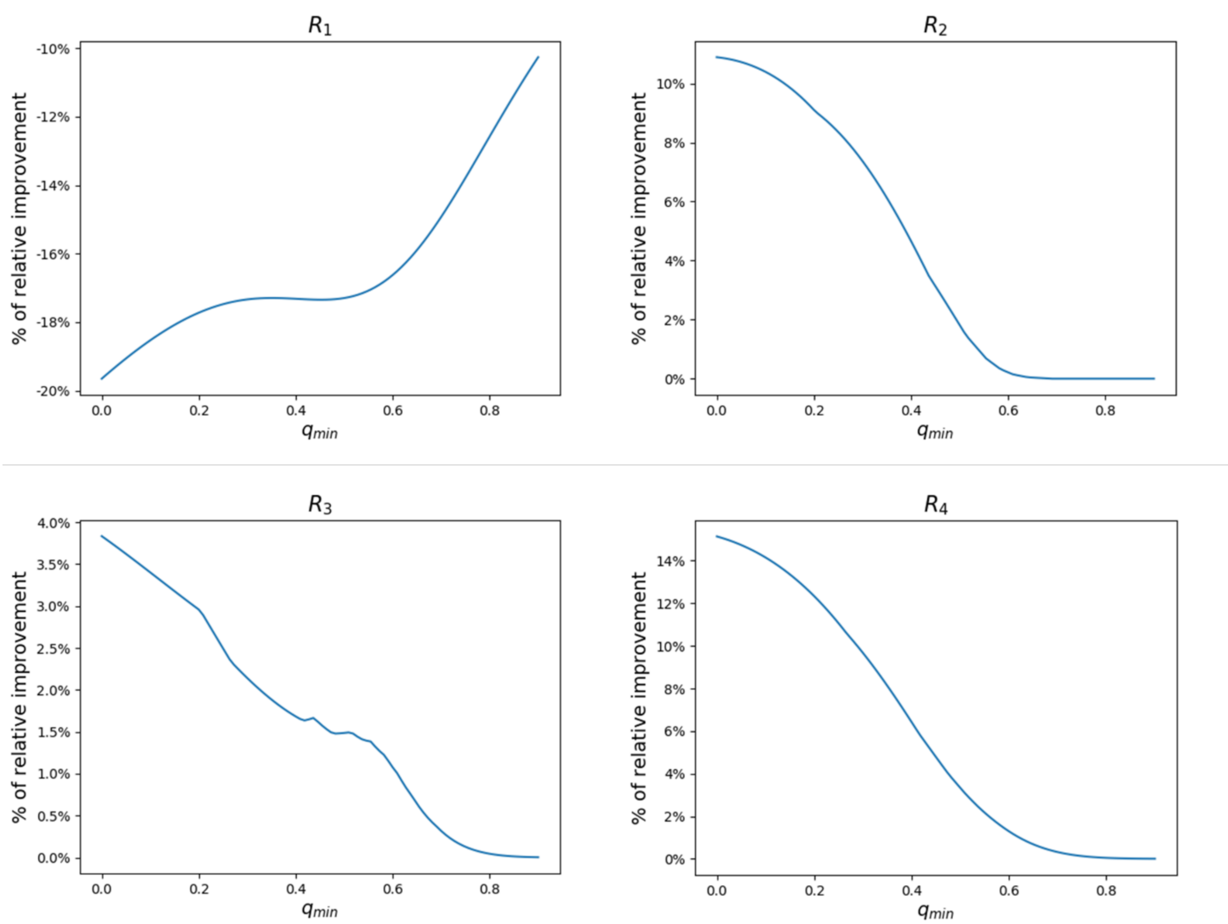

Figure B.3. Measures of improvements as q_{min} varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$.

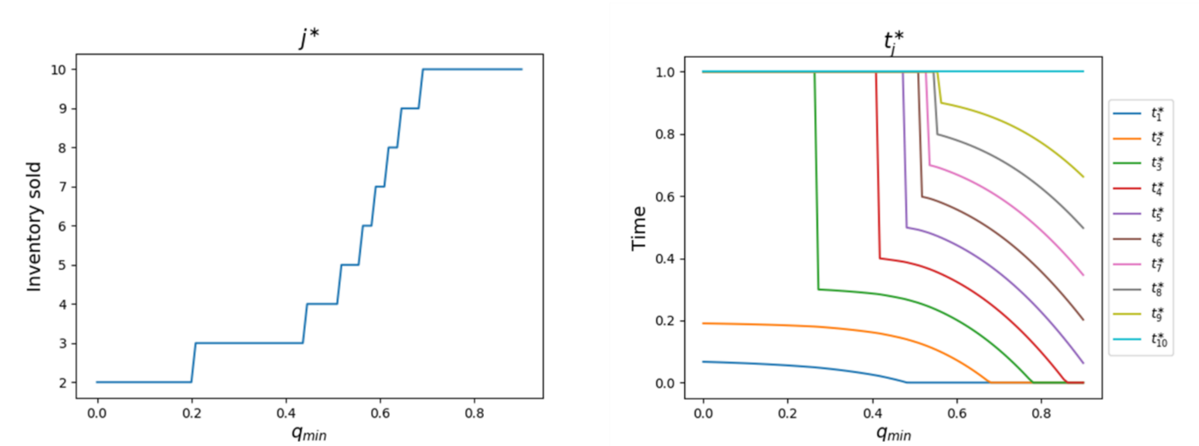

Figure B.4. The optimal thresholds of the fixed (left) and time-dependent (right) threshold policies as q_{min} varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\sigma_\varepsilon = 0.25$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$.

Figures B.5 and B.6 illustrate the impacts of σ_ε on the four measures of improvements and the associated thresholds. We find that R_1 is first increasing then slightly decreasing in σ_ε, R_2 and R_4 are unimodal with respect to σ_ε, and R_3 is decreasing in σ_ε. When σ_ε is small, as most customers perceive that the product is not of high quality under our default parameter setting, they would not buy it if inventory information is disclosed and the customers' purchase decisions are influenced by observational learning. As σ_ε grows, the random factor in customer's propensity of purchase increases, and customers are more likely to purchase the product, which contributes to the increase of R_1 and R_2. However, when σ_ε is large, the relative advantages of the fixed and time-dependent threshold policies compared to the better of the two simple policies diminish as the expected sales of both simple policies increase as well. The decreasing pattern of R_3 is due to the fact that when σ_ε is higher, the randomness of the customers' purchase decisions increases, and thus, the relative improvement of the time-dependent threshold policy compared to the fixed threshold policy becomes less significant. The behavior of R_4 is due to the combined effect of both

R_2 and R_3. Furthermore, as Figure B.6 shows, the optimal fixed threshold decreases and the optimal time thresholds increase in σ_ε, both encouraging the disclosure of inventory.

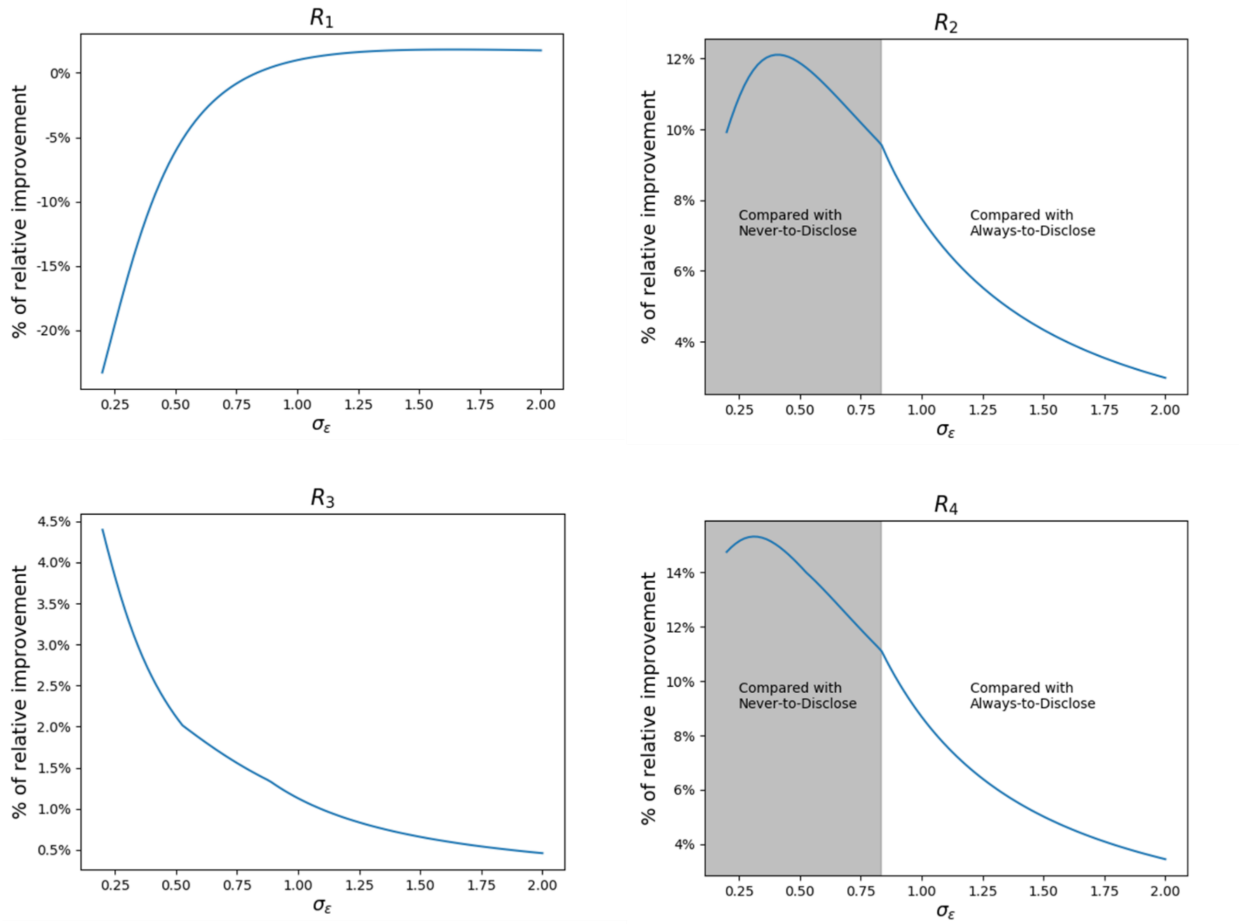

Figure B.5. Measures of improvements as σ_ε varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

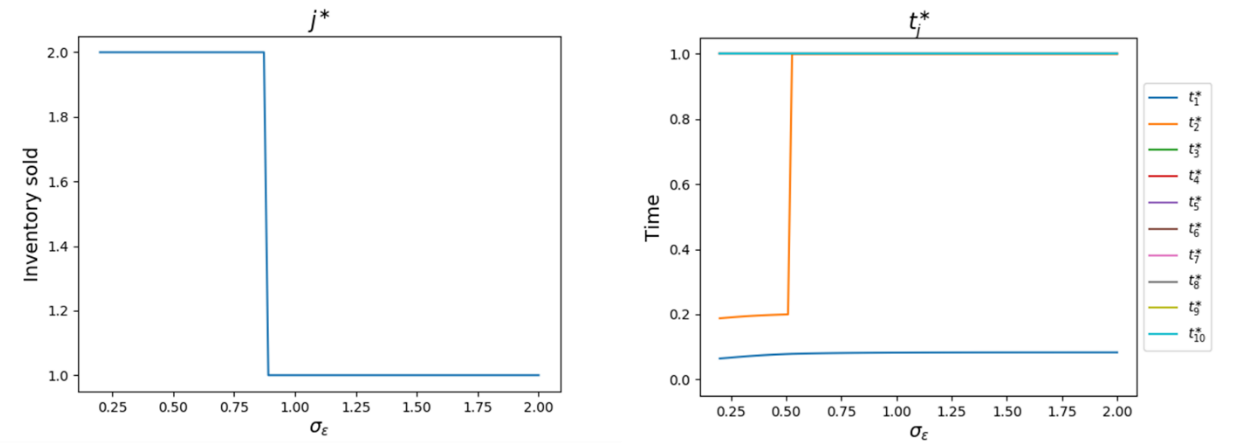

Figure B.6. The optimal thresholds of the fixed (left) and time-dependent (right) threshold policies as σ_ε varies.

Note. $b_1 = 1.2$, $b_2 = 0.8$, $\rho_{max} = 0.8$, $p_{so}^{ND} = 0.3$, $p_{so}^{BD} = 0.3$, $\Lambda_0 = 10$, $a = 0$, $K = 10$, $T = 1$.

VITA

Zibo Liu is a researcher in the field of Information Systems. His research interest lies in promotional strategies on online platforms such as e-commerce platforms and crowdfunding platforms, product sampling, platform designs, and various mechanisms on online platforms. His research methodologies include structural modeling, econometrics, economic theories, machine learning and AI, queueing theory, inventory analysis, and analytical models in IS. He receives his Ph.D. degree at the Michael G. Foster School of Business, University of Washington. Prior to joining UW, he completed his undergraduate study at Department of Electronic Engineering of Tsinghua University. He will be joining School of Management at Fudan University as an Assistant Professor as of August 2022.