ESSAYS ON TECHNOLOGY-DRIVEN MARKETING

Zijun (June) Shi

Submitted to the Tepper School of Business in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Dissertation Committee

Kannan Srinivasan (Chair)
Hui Li
Xiao Liu
Alan Montgomery
Kaifu Zhang

April, 2019

© Coypyright by Zijun (June) Shi, 2019
All rights reserved.

Abstract:

With the development of technology in business applications, new marketing problems emerge, creating challenges for both practitioners and researchers. In this dissertation, I investigate marketing issues that involve new technology or require research methodologies enabled by new technology. I take an interdisciplinary approach, combining structural modeling, analytical modeling, machine learning, and causal inference, to study problems on pricing, media hype, and branding in three essays.

In the first essay, we examine the optimality of the freemium pricing strategy. Despite its immense popularity, the freemium business model remains a complex strategy to master and often a topic of heated debate. Adopting a generalized version of the screening framework à la Mussa and Rosen (1978), we ask when and why a firm should endogenously offer a zero price on its low-end product when users' product usages generate network externalities on each other. Our analysis indicates freemium can only emerge if the high- and low-end products provide asymmetric marginal network effects. In other words, the firm would set a zero price for its low-end product only if the high-end product provided larger utility gain from an expansion of the firm's user base. In contrast to conventional beliefs, a firm pursuing the freemium strategy might increase the baseline quality on its low-end product above the “efficient” level, which seemingly reduces differentiation.

In the second essay, we study how misleading health information affects subsequent information generation from different sources, by examining publicly available health information about over-the-counter (OTC) weight loss products. We find that biased health information from celebrity doctors may initiate a media-hype, creating news waves on the recommended healthcare products. By analyzing the news content using natural language processing techniques, we find that the news articles generated after The Dr. Oz Show carry higher sentiment and more positive emotion, and they provide little correction for the hype news. While the government has been fighting against fake news in the healthcare domain, our finding suggests that even legitimate and genuine news articles respond to the hype news by propagating and magnifying the hype news, rather than correcting them. Research articles either react too slowly or do not respond to biased information. Though some user-generated content (UGC) provides correction, it is overwhelmed by the larger amount of UGC that supports the misleading information. Lastly, public information remained biased until the Federal Trade Commission hearing and investigation. Our findings have public policy implications on how to potentially regulate media content in the healthcare domain to protect consumers from misleading health information.

In the third essay, we develop a dynamic structural model of fashion choices of brands and styles to investigate the implication of prohibiting fast fashion copycats (e.g., Zara, Forever 21), with the help of user-generated data from fashion-specific social media and deep learning methods on image analytics. We find that copycats can enhance high-end brands demand, contrary to

conventional wisdom, due to several novel mechanisms: first, the affordability of mixing low-end copycats with high-end brands boosts demand for high-end brands from financially constrained consumers; second, good styles from low-end brands can help a consumer to build up his popularity/likeability, which increases his value for high-end brands and reduces the cost. Substantively, our results shed light on copyright enforcement and have implications on how fashion brands should react to copycats. Methodically, we developed a framework to analyze consumer choices where visual features are important product attributes and peer feedback hugely affects the decision-making process.

Acknowledgment:

I am deeply indebted to the support and guidance of my dissertation committee members: Professor Kannan Srinivasan, Professor Kaifu Zhang, Professor Xiao Liu, Professor Alan Montgomery, and Professor Hui Li. Since my first sememster in Ph.D. program, I was constantly enlightened by the discussion with Professor Kannan Srinivasan. Throughout the writing of this dissertation, I have received a great deal of support and utmost encouragement from him. Professor Kaifu Zhang guided me on conducting analytical research in Marketing, with insightful direction and invaluable advice. His expertise guided me especially on formulating the research topic and methodology. Professor Xiao Liu provided me with unreserved support since I moved on to the empirical research domain. I benefited a lot from every discussion with her. I also want to express my gratitude to Professor Alan Montgomery and Professor Hui Li. Their expertise and knowledge make the process of developing this dissertation rewarding and exciting.

I would also like to thank Professor Param V. Singh. His invaluable advice and support are tremendously helpful for my research and the job market. I learned not only research skills but also precious experience from him. I would also like to thank other professors: Dokyun Lee, Christopher Y. Olivola, Peter Boatwright, Jeff Galak, and Tong (Joy) Lu, for taking the time and effort discussing research ideas and giving advice. Their support makes the past five years a more rewarding and pleasant journey.

I thank my husband Siyu Lu. This dissertation would not have been possible without his endless support, unreserved love, and continued patience. I would also like to express my deepest gratitude to our parents for their love and support. Lastly, I thank my daughter, Yutong Thea Lu, for the great joy she brought to my life and for the distraction she injected into my work time.

For my dearest family

Contents

Chapter 1

Freemium as an Optimal Strategy for Market Dominant Firms

1.1 Introduction

Over the past years, "freemium" has attracted considerable attention from both practitioners and academics. Many believe that freemium underlies the meteoric rise of companies like Skype and Dropbox, and a horde of startups have jumped on the bandwagon and adopted freemium as their business model of choice. However, successful implementation of the freemium strategy remains challenging. A WSJ report titled "When Freemium Fails" interviewed frustrated entrepreneurs who considered "move(ing) away from freemium" as "the best business decision ... (they) ever made" (Needleman and Loten 2012). An investment manager at First Round Capital summarized the entrepreneurs' frustration as "too many freemium models have too much free and not enough mium". From online gaming to music streaming, leading companies offer widely divergent opinions on whether and when a free option should be offered at all.[1]

Not only is freemium controversial among practitioners, but it also represents a curious case in the eyes of a theoretician. On the surface, freemium resembles a classic case of product-line screening, wherein a firm offers a menu of products at different prices to segment the market. However, as proven by Mussa and Rosen (1978) and more recently by Anderson and Celik (2014), a profit-maximizing firm should always choose inefficient quality

[1]See Zetlin (2013), for example, an interview of Rhapsody's CEO who insisted on its subscription-only model while competitors adopted freemium. In the gaming industry, leading firms such as Blizzard Inc. offer freemium on some of their games but not others.

but efficient price for its lowest quality product, while doing the reverse for its highest quality product.[2] Said differently, the low-end product's price should be positive and maximize the single-product profit. This theoretical prescription seems to stand in exact opposite to the notion of freemium.

A number of straightforward reasons come to mind as to why firms find freemium attractive. The illusion of advertising as a "last resort" revenue source is often cited. Saving users the hassle of payment (which can seem high when the price is too small) is another. The power of "free" as a behavioral marketing tool is a third. Although these factors are certainly relevant, we join a nascent literature in marketing and information system that looks at the more nuanced economic reasons behind the freemium phenomenon. In marketing, two pioneering papers by Kamada and Ory (2015) and Lee et al. (2013) have studied the design of freemium to facilitate word-of-mouth and product diffusion. In comparison, we adopt a single-period monopolistic screening framework and study the optimality of freemium when diffusion-related factors are absent. That is, we ask whether and when "perpetual freemium" remains an effective strategy once a product has achieved sufficient recognition and diffusion-related factors have declined in importance.[3] This is especially relevant for firms that have almost reached market saturation. Google Drive and LinkedIn are best examples where word of mouth or diffusion is a non-issue.

More specifically, we ask whether network effects from product usage alone can justify the freemium model, when a firm's sole objective is its single-period product line profit. The notion of network effects speaks to the fact that consumers' valuation of a product varies depending on how many other consumers are using the product or compatible products. Network effects can be generated not only by direct interactions of consumers, but also by indirect behavioral reasons. Direct interaction happens when a free user of Dropbox shares a file with a paid user, when a free player of Farmville trades with a paid user, and when

[2]Throughout the paper, we use the term efficient quality to refer to the quality level that maximizes single product profit (therefore social welfare) under the complete information benchmark.

[3]A number of papers have touched upon this issue (Niculescu and Wu, 2014; Cheng and Tang, 2010), but, none of them have completely endogenized prices and qualities in the product line.

a free user of Spotify shares her playlist. Behaviorally, network effects are created when a consumer values the product more if there are more users of the same product because it allows him/her to socially fit in with their peers, or when a consumer values the paid product more if there are more users of the free product because he/she can derive social prestige from using the high-end product. Intuitively, offering a free version brings more users on board and generates greater network effects. Meanwhile, the Mussa and Rosen (1978)'s insights remain valid and the risk of cannibalization remains high. Ex-ante, it is not clear which is the dominant factor.

By endogenizing a firm's price decision in all relevant subgames, in the baseline model, we build a general framework to study a firm's product line strategy. We pay particular attention to whether and when the firm would *endogenously* choose a zero price for its low-end product. Our first set of results speaks to the conditions under which freemium will *not* hold. We show that, as expected, freemium cannot emerge in a classic screening model without network effects. We are able to prove this with a very general quasi-linear utility function and type distribution, thereby showing the robustness of the Mussa and Rosen (1978) insights. Importantly, even with uniform network effects, freemium remains a dominated strategy. Although network effects lead to stronger incentives to expand the market, they also make the cannibalization effect stronger as more users adopt the low-end ("free") product. When the free product delivers the same network value as the paid product, a price cut on the low-end product will increase the attractiveness of both products by the same margin, thus *tightening* each consumer's incentive compatibility constraint. Thus, more free users does not translate into higher profits from the paid users. Consequently, although network effects give the firm stronger incentives to cover the market, this is done by offering a "conventional product line," wherein the low quality product is priced at a positive level. This result remains valid when we endogenize quality choices for the product line. Put differently, introducing sufficient difference only in "standalone" qualities is not enough to make the freemium strategy viable.

We further show that freemium could indeed become optimal when there is sufficient asymmetry in network effects between the high- and low-end products. In order for freemium to be viable, the firm's product line has to be such that the paid users gain access to larger network effects compared to their non-paying peers. This result somewhat echoes the message in Kumar (2014) that in order to make freemium work, a firm has to offer different sets of features in its free and paid products. But we show that it is the network effects, rather than the "standalone" quality that are the crucial factors.

As an extension, we endogenize the firm's quality decision of products and examine how the optimal quality levels should change with respect to the network effects. To do so, we consider a specific linear utility function and a uniform distribution for consumer type. We show that, in conventional product line design, the optimal quality of the low-end product increases when network effects are higher. Put differently, standalone quality and network effects are complementary to each other. However, in a freemium equilibium, the low-end product's quality decreases when network effects are larger. In other words, low-end product's network effect is a substitute to its standalone quality. This result stems from the fact that in a freemium equilibrium, the entry-level product generates no revenue and its own purpose is to expand a firm's user base. The quality provision should be "just enough" to bring enough users on board.

As a second extension, we fully endogenize the firm's product line decisions – quality, price, as well as network effects – by examining a simpler model where type distribution is discrete. Remarkably, the main insights from the general model remain intact. We compare the qualities of both products provided under freemium and those offered under conventional product line design. Our analysis suggests that, when the firm adopts the freemium strategy, the (non-network) quality gap between the high-end and the low-end products actually shrinks. When adopting freemium strategy, the firm should even provide a low-end product whose quality is above the efficient level. In other words, the firm should offer higher quality and a zero price in order to retain the low-type consumers; this surprising result stands in contrast to the Mussa and Rosen (1978)'s results of efficient price and inefficient quality. In an optimal product line, quality and network effects are substitutes for the low-end product, but are independent dimensions for the high-end product.

To sum up, our analysis yields a set of managerial recommendations that complement what has been suggested in the previous literature. We show that in the absence of word-of-mouth, "getting more consumers on board" alone cannot justify the freemium strategy. In most cases, market expansion can be more effectively achieved by offering a conventional product line, wherein the low-end product is priced at a positive level to avoid unnecessary cannibalization. In the current framework, perpetual freemium is only optimal under network effects asymmetry. The right freemium strategy should include a free product with lower network benefit than the paid product, but superior standalone functionalities (compared to the efficient level).

The difficulty of establishing freemium arising in equilibrium needs special mention. When consumers derive positive utility for a product, the firm can always charge a positive price and such a deviation is inherently likely to be more profitable. Thus, in general, sustaining freemium in equilibrium is likely to be difficult. Given the extensive and growing nature of freemium, however, demonstrating that such a strategy can arise in equilibrium assumes significant importance.

In the markets where freemium is common, externality benefits are also quite common. Therefore, incorporating externality is a natural and arguably critical element of model formulation to examine product line price and qualities in equilibrium. In spite of a rich model, we get a sharp insight that asymmetric network externality is essential to support freemium. Equally important sharp insight is that freemium will not be sustained when the network effects are the same across all levels of products that differ in quality.

We organize the rest of the paper as follows. In Section 1.2, the related literature and the contribution of the present paper are discussed. Section 1.3 presents the model setup. Section 1.4 presents the analysis and results. Section 1.5 discusses the extension on endogenous quality and network effect decisions, as well as the discrete case. Section 1.6 concludes.

1.2 Literature Review

This paper is related to three streams of literature. First, a number of recent papers have studied various aspects of the freemium strategy. In marketing, Lee et al. (2013) and Kamada and Ory (2015) are among the first studies on the design of freemium. Kamada and Ory (2015) build a micro model of referral behavior and investigate whether a free contract or a referral program is a more efficient means to encourage word of mouth (WOM). The free contract ensures that a receiver would adopt the product even if she turns out to be a low type. When a receiver's adoption generates network effects on the sender, the free contract gives the sender stronger incentives to refer the product in the first place. The main trade-off is between expanded second period demand (due to WOM) and cannibalization. Our model shares some of the features in Kamada and Ory (2015). However, instead of network externalities between first-period senders and second-period receivers, we consider a static model in which network effects exist within and between consumer segments. In our model, an expanded network size leads to the potential to increase the high-end product's price, while a zero price for the low-end product leads to cannibalization. In other words, the focus of this paper is on the optimality of freemium when diffusion dynamics are absent (Mahajan et al., 1990; Chatterjee and Eliashberg, 1990).

In another closely related paper, Lee et al. (2013) develop a structural model to study the design of freemium. Although the paper's focus is empirical, it develops a rich model of consumer behavior that encompasses adoption, upgrade, referral and usage. There are two main differences between Lee et al. (2013) and this paper. First, this paper considers network effects from product usage but does not model diffusion dynamics. Second, Lee et al. (2013) study the design of freemium once the firm has already committed to a zero price for its low-end product. Even though this is a realistic setup in many contexts wherein a firm would commit to freemium for strategic reasons,[4] we are interested primarily in when and why freemium would endogenously emerge to maximize product line profit. Thus, we endogenize the price on the low-end product instead of fixing it to zero.

A number of papers from information systems have studied various aspects of free trial, popular in the software industry (Cheng and Tang, 2010; Niculescu and Wu, 2014; Cheng and Liu, 2012). None of these papers have fully endogenized price in a general model with a general distribution of consumer type. In particular, we allow the low-end product's price to be endogenous.

Therefore, our study is closely related to the rich literature on product line design in both economics and marketing (e.g., Mussa and Rosen 1978; Anderson and Celik 2014; Desai 2001; Desai et al. 2001). We follow the paradigm established in Mussa and Rosen (1978) and consider single-period product line profit as the firm's objective function. The firm chooses how many products to offer and sets a price for each product. As shown more recently in Anderson and Celik (2014), without network effects, the optimal product line problem can be reformulated as a multi-step optimization problem. The firm first chooses the lowest-quality product's price to maximize its revenue, then proceeds to maximize the additional revenue that comes from the second-lowest-quality product. While the standard Mussa and Rosen model does not consider network effects, a number of recent papers have examined the impact of network effects. Jing (2007) examined market segmentation under network externalities and found that the existence of network effects gives the firm stronger incentives to cover the market. The author did not explore the case of freemium, but his main insights are echoed in this paper.

Finally, in a broad sense, an asymmetric product line is somewhat reminiscent of a two-sided market. In a two-sided market setup, a platform (firm) has incentives to lower the price for one side below the marginal cost as long as this brings value to the other side (e.g.,

[4]For example, keeping a product for free saves the need to set up a payment system, potentially lowering a firm's operating cost. Similarly, zero price has been shown to be a particularly powerful marketing tool, and might be preferred to a small but positive price for behavioral reasons.

Hagiu 2006; Rochet and Tirole 2006). Although cannibalization is not relevant in a two-sided market context, it is in the context of product line design. We show that it is the coupling of network effects and cannibalization effect that makes the freemium problem unique.

1.3 Model

In analyzing the freemium problem, we intend to make our key insights as general as possible and independent of most specific assumptions on functional form. Thus, we start by presenting a general model in which we make few assumptions on the form of the utility function as well as consumer type distribution. At the same time, we present a running example with linear utility function and uniform distribution of consumer type. This allows us to precisely pin down the conditions for freemium in analytical forms, and we hope that this exercise will strengthen our main intuitions. Next, we present the general model and the running example in turn.

Consider a monopolist who has the option of offering either one or two vertically differentiated products. For notational convenience, we denote the firm's product strategy as Ω. If two products are offered, $\Omega = \{L, H\}$, where L, H stands for the products of relatively low and high quality. If only one product is offered, Ω is a singleton.

There is a unit mass of consumers. They have heterogeneous taste, which is described by the distribution of θ, a density $f(\theta)$ defined on $[0, 1]$. All customers have access to an outside option, the utility of which is u_0. In the case of Dropbox, for example, this outside option denotes the utility a consumer gets from using a traditional form of storage.[5]

For a customer with taste parameter θ, her valuation from consuming product i is $V^i(\theta, D)$, where $i \in \Omega$ and D is the total user base of the firm's product. Let $D = D^{-i} + D^i$ where D^i is the demand for product i, and D^{-i} is the demand for the other type of product if offered. In this framework, $\frac{\partial V^i(\theta,D)}{\partial D}$ captures network benefit for consumers using product i.

[5]In the main text, we consider cases where outside options for all consumers are the same. In Online Appendix B, we provide analysis for the cases where consumers have heterogeneous outside options. In short, heterogeneous outside options do not qualitatively affect our results in the main analysis.

In other words, we consider a type of "global" network effects where only the total network size affects consumer utility, though the relationship does not have to be linear. The total utility a consumer derives from buying product i is therefore $U^i = V^i(\theta, D) - p^i$, where p^i is the price of product i.

We make the following assumptions regarding $V^i(\theta, D)$.

AI. strictly increasing in quality. $\forall \theta, D$,

$$V^H(\theta, D) > V^L(\theta, D)$$

AII. differentiable in D, *and* strictly increasing in D. $\forall \theta, i$,

$$\frac{\partial V^i(\theta, D)}{\partial D} > 0$$

AIII. differentiable in θ, *and* strictly increasing in θ. $\forall D, i$,

$$\frac{\partial V^i(\theta, D)}{\partial \theta} > 0$$

AIV. has increasing differences in θ and quality/network effects. $\forall D$,

$$\frac{\partial\left[V^H(\theta, D) - V^L(\theta, D)\right]}{\partial \theta} > 0$$

The assumption *AIV* can be alternatively and more restrictively stated as two assumptions that, respectively, speak to the increasing differences conditions regarding consumer type and the standalone product quality/ network effects, i.e., $\frac{\partial\left[V^H(\theta,0) - V^L(\theta,0)\right]}{\partial \theta} > 0$ and $\forall D$, $\frac{\partial\left[V^H(\theta,D) - V^L(\theta,D)\right]}{\partial D \partial \theta} \geq 0$[6]. Here, $V^i(\theta, 0)$ captures the consumer's valuation of product i's quality when consumers do not derive utility from other users (equivalently $D = 0$), and $\frac{\partial V^i(\theta,D)}{\partial D}$ captures the marginal network effect derived from using product i which has user base D.

[6] These two requirements together are sufficient conditions of assumption *AIV*, please see appendix for proof.

The first inequality implies $V^i(\theta, 0)$ has increasing difference in (θ, i), while the second inequality means the network benefit also has increasing difference in (θ, i). Assumption *AIV* relaxes the conditions, and requires only $V^i(\theta, D)$ has increasing difference in (θ, i).

The game consists of two stages. First, the firm chooses its menu of products. Next, consumers make purchase decisions conditional upon the firm's menu and belief about all other consumers' decisions. As is typical in a game with network effects, multiple equilibria may exist in the second stage. Specifically, we seek the Nash Equilibrium in the second stage game that is Pareto dominant. Assumption *AIII* guarantees that a consumer of type θ_0 would expect that all other consumers with $\theta > \theta_0$ have a higher evaluation for any product. Therefore, if type θ_0 prefers purchasing product i to the outside option, type $\theta > \theta_0$ would also do so. Similarly, if type θ_0 prefers purchasing the higher-quality product to the lower-quality one, he would expect type $\theta > \theta_0$ to do the same. In the appendix, we show that in the baseline model, the Pareto dominant outcome consists of one of two outcomes. In the first scenario, the firm offers the high-quality product only. In the second scenario, the firm offers a product line, with the higher-valuation segment buying the high-quality product and the lower-valuation segment buying the low-quality product.

The assumptions made above are consistent with those in classic papers on product line design, except we also account for network effects. Lemma 1 illustrates how the demand schedules are determined for each product and pricing strategy.

Lemma 1. *Assume that both types of products are offered, and the prices are such that both have positive sales. The demand schedule is determined by two marginal consumers at locations θ_L and θ_{HL}, where*

$$V^L(\theta_L, D) - p^L = u_0 \tag{1}$$

$$V^H(\theta_{HL}, D) - p^H = V^L(\theta_{HL}, D) - p^L \tag{2}$$

with the low-end product serving $[\theta_L, \theta_{HL}]$, and the high-end product serving $[\theta_{HL}, 1]$.

When only the high-end product is offered, the marginal customer type θ_H is determined by

$$V^H(\theta_H, D) - p^H = u_0 \tag{3}$$

and only consumers with $[\theta_H, 1]$ are served.

The proof for Lemma 1 is as follows. In choosing between two types of products, a consumer of type θ chooses the low-end product only if $p^H - p^L > V^H(\theta, D) - V^L(\theta, D)$. Given assumption AIV, the larger θ is, the greater the reduction in price is required for a consumer to choose the low-end product. Hence, it is impossible to induce a consumer of type θ_i to purchase a low-end item if the high-end product is purchased by a consumer of type $\theta_j < \theta_i$. From this feature of $V^i(\theta, D)$ and from the fact that the monopolist can make positive profits from serving at least the high-θ consumers, it follows that if both types of products are offered, the monopolist serves $[\theta_L, \theta_{HL}]$ with the low-quality product, and $[\theta_{HL}, 1]$ with the high-quality product, where θ_L denotes the marginal consumer who is indifferent between purchasing the low-end product and the outside option, while θ_{HL} denotes the marginal consumer who is indifferent between purchasing the high-end and low-end product. The firm's profit is thus $\Pi_{HL} = D^H p^H + D^L p^L$. In the case where only the high-end product is offered,[7] the firm serves $[\theta_H, 1]$, where θ_H denotes the marginal customer who is indifferent between purchasing the high-end product and the outside option. In this case, the firm's profit is simply $\Pi_H = D^H p^H$.

A freemium equilibrium is one in which the firm offers both types of products but charges a zero price for the low-end product. It is formally defined as follows:

Definition 1. An freemium equilibrium is defined as a product line offering both products H and L, wherein $p^{L*} = 0,\ p^{H*} > 0,\ \Pi^*_{HL} > \Pi^*_H$.

[7]Note that it is not an equilibrium wherein the firm sells only the low-end product. Comparatively, offering only the low-end product is dominated and trivial, because the firm can always attract the same user bases with higher price by offering the high-end product. Therefore, selling only the low-end product is trivially dominated.

In order to explain our findings more clearly, throughout the paper, we will illustrate our general findings with a running example with valuation function $V^i(\theta, D) = \theta q^i + \alpha^i D$ and uniform distribution of consumer types, i.e., $\theta \sim U[0,1]$. The quality of product i is denoted by q^i, and α^i captures the network benefit one can derive from any other user's usage of product i. This running example satisfies all the assumptions $AI \sim AIV$.

In this running example, the demand for each product and the total demand are given by

$$D^H = \int_{\theta_{HL}}^{1} f(\theta)\, d\theta = 1 - \theta_{HL},$$

$$D^L = \int_{\theta_L}^{\theta_{HL}} f(\theta)\, d\theta = \theta_{HL} - \theta_L,$$

$$D = D^H + D^L = 1 - \theta_L,$$

where the marginal consumers θ_L, θ_{HL} are given by

$$\theta_L q^L + \alpha^L D - p^L = u_0,$$

$$\theta_{HL} q^H + \alpha^H D - p^H = \theta_{HL} q^L + \alpha^L D - p^L.$$

The total profit for the monopolist is thus

$$\Pi_{HL} = p^H D^H + p^L D^L.$$

In the baseline analysis, we therefore endogenize the product set choice as well as the pricing decision. We do not, however, endogenize the quality level nor the level of network effects. In extensions, we endogenize both decisions by considering model formulations that are analytically tractable. According to Definition 1, freemium is different from selecting a price equal to the marginal cost. As demonstrated in extensions (Section 1.5) where quality decision is endogenized, incorporating a positive production cost does not qualitatively affect our results in the main analysis (Section 1.4).

Before we proceed with the analysis, let us briefly discuss our formulation of network effects. We choose a relatively simple formulation wherein each product delivers a different level of network effect (e.g., α_L and α_H in the running example). This is a somewhat standard formulation where network effect is considered as a product attribute. In reality, however, the patterns of network effects can be much more complex. For example, in social games, the network effects are governed by the consumers' obtained utilities when they interact with each other in games. The non-paying users may play at a disadvantage and derive a disutility when interacting with paying users. Though this scenario will indeed imply that the free option generates less overall network effects than the paid option, the exact level of disutility versus utility would depend on the frequency of interaction between the two types of users. In other words, it can be best captured by a case of *local* network effects with four, instead of two, parameters. Throughout the analysis, we choose to present the simplest model where the network effects can be parameterized by two parameters. It should be kept in mind, that our notion of asymmetric network effects can be richer than it seems, capturing cases such as the social game example discussed above. A more detailed analysis that covers the case of quality-dependent and local network effects is in Online Appendix A.

1.4 Analysis

The game has two stages: first, the firm decides whether to offer one or two products, and sets the price for the offered product(s); second, consumers decide on whether to purchase the product that gives the highest utility, or purchase nothing. To explore the conditions under which freemium is an optimal strategy, we discuss three cases in turn:

(1)No network effect $\frac{\partial V^i(\theta,D)}{\partial D} = 0$ for all $i \in \{L, H\}$.

(2)Uniform (symmetric) network effect $\frac{\partial V^i(\theta,D)}{\partial D} = \alpha > 0$ for all $i \in \{L, H\}$, where α is a constant.

(3)Asymmetric network effect $\frac{\partial V^i(\theta,D)}{\partial D}$ differs for $i = L$ and $i = H$.

To prove the optimality of freemium, we consider a necessary condition for freemium: $\frac{\partial \Pi}{\partial p^L}|_{p^L=0} < 0$. In words, the monopoly would like to set a zero price for the low-end product, only when he has no incentives to marginally increase the price when it is already at zero. It turns out that this condition is sufficient to rule out the optimality of freemium in cases (1) and (2), under the general functional form. Let θ_L,θ_H and θ_{HL} be defined by equations $(1), (2), (3)$, and the asterisk $*$ refer to the equilibrium value. Assuming uniform or no network effects, Proposition 1 states the non-existence of freemium under uniform network effects while Corollary 1 illustrates it with the running example.When $\frac{\partial V^i(\theta,D)}{\partial D} = \alpha \geq 0$ for $i \in \{L, H\}$, freemium can never emerge as the equilibrium strategy.

Proposition 1. *When* $\frac{\partial V^i(\theta,D)}{\partial D} = \alpha \geq 0$ *for* $i \in \{L, H\}$*, freemium can never emerge as the equilibrium strategy.*

For the running example, Proposition 1 can be stated in a more explicite way as in Corollary 1.

Corollary 1. *With* $\alpha^i = \alpha \geq 0$ *for* $i \in \{L, H\}$*,* $V^i(\theta, D) = \theta q^i + \alpha D$ *and* $\theta \in U[0, 1]$*, the firm's equilibrium product and pricing strategy can be characterized as follows :*

(a) When $q^L > \alpha > u_0$ *and* $q^L + u_0 \geq 2\alpha$*, the firm offers both products with* $p^{L*} = \frac{q^L - u_0}{2}$*,* $p^{H*} = \frac{q^H - u_0}{2}$*,* $\theta^*_{HL} = \frac{1}{2}$*,* $\theta^*_L = \frac{q^L + u_0 - 2\alpha}{2(q^L - \alpha)}$ *and* $\Pi^*_{HL} = \frac{q^L - u_0}{4}\left(\frac{\alpha - u_0}{q^L - \alpha}\right) + \frac{q^H - u_0}{4}$*.*

(b) Otherwise, the firm offers only the high-end product, with $p^{H*} = \frac{q^H - u_0}{2}$*,* $\theta^*_H = \frac{u_0 + q^H - 2\alpha}{2(q^H - \alpha)}$ *and* $\Pi^*_H = \frac{\left(q^H - u_0\right)^2}{4(q^H - \alpha)}$*.*

We explain the intuition behind Proposition 1 with two sub-cases: one without network effects (i.e., $\frac{\partial V^i(\theta,D)}{\partial D} = 0$ for all $i \in \{L, H\}$), and the other with uniform network effects (i.e., $\frac{\partial V^i(\theta,D)}{\partial D} = \alpha > 0$ for $i \in \{L, H\}$). Although the former is a special case of the latter, separating them helps us build some intuitions. The intuition for the case with zero network effects is consistent with what has been shown in the product line design literature, and has been most succinctly summarized by Anderson and Celik (2014). In a nutshell, setting the low-end product's price to zero generates no revenue *and* puts downward pressure on the high-quality product's price and demand. Thus, the firm can be better off by withdrawing the low-end product altogether. When it does lauch the entry-level product, it always sets a price that is efficient from a single product profit maximization standpoint (see Anderson and Celik 2014 for details).

When network effects are present, how would the firm's optimal product strategy be impacted? More specifically, can network effects lead to an equilibrium wherein the firm pursues the freemium strategy? As discussed in the introduction, increasing network size is a major intuition in favor of the freemium strategy. Although the low-end product generates no revenue and partially cannibalizes the high-end sales, a larger network size brings higher utility to the high-valuation customers, possibly leading to higher price and, therefore, profit. This is akin to the strategy of user subsidization in a two-sided market context. However, a formal analysis reveals that this intuition is not valid when the network benefits for users of both high- and low-end products are positive but identical.

What is the intuition behind? For freemium to be optimal, the optimal price for the low-end product must be zero. A necessary condition for this is $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^L=0} \leq 0$. In other words, the firm has incentive to decrease the low-end price even it is already at or close to zero. When network effects are symmetric or not too different between the two products, this condition cannot be met. At $p^L = 0$, it is straightforward that $\frac{\partial \Pi_L}{\partial p^L}|_{p^L=0} \geq 0$. Thus, $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^L=0} \leq 0$ requires that $\frac{\partial \Pi_H}{\partial p^L}|_{p^L=0} \geq 0$. In other words, at an infinitesimal p^L, the profit from the high-end product would increase as the firm decreases its price on the low-end product. A necessary condition for this is that the marginal consumer's incentive compatibility constraint is relaxed as p^L is reduced. This is necessary for either the demand or the price of the high-end product to increase. However, under uniform network effect, this is not possible for the following intuition. As the firm lowers the price for its low-end product, the total user base expands. Due to network effects, each user indeed gains greater surplus from using the high end product. However, because network effects are symmetric, the low-end product also becomes more attractive, by the same margin. In other words, each consumer's incentive compatibility constraint has not been relaxed. The marginal consumer, in fact, now prefers the low-end product more because of a lower price. Thus, the firm cannot charge a higher p^H nor will it have a higher demand from the high-end product. Meanwhile, the firm suffers greater loss from the low valuation segment. Taken together, in the subgame wherein the firm offers two products, decreasing p^L while it is close to zero, always decreases firm profit. Please see the appendix for the formal proof.

In practice, it is certainly rare that different products would deliver exactly the same level of network effects. However, the broad insights remain intact as long as the network benefits of different products are close enough. This corresponds to a wide range of applications where the firm does not or cannot restrict interaction between paid and free users. In most mobile messaging tools, for example, users can send messages to each other regardless of whether they are paying or not. The network aspect of the product is a relatively simple and straightforward feature. It is difficult to restrict the network benefits enjoyed by the non-paying users unless the firm intentionally handicaps the product.

For the case of uniform network effect, Proposition 1 uncovers a fundamental tension between expanding the network size and containing cannibalization. Next, we consider the case in which the firm's high-end and low-end products can deliver different levels of network effects. This is a widely observed practice among firms who successfully pursue the freemium strategy. LinkedIn, for example, gives free users only limited access to view others' profiles, especially contact details. In some of its freemium games, Zynga used to charge "entry tickets" if the users wanted to game with other users. Paid users of Dropbox are able to share more files with others than free users. Proposition 2 states that freemium may indeed emerge as an equilibrium strategy when the high-end and low-end products differ on both the baseline quality as well as the network effects dimension.

Proposition 2. *When the following necessary condition is satisfied, freemium can be an* *equilibrium strategy:*

$$\left[\frac{\partial V^H(\theta_{HL}, D)}{\partial D} - \frac{\partial V^L(\theta_{HL}, D)}{\partial D}\right]|_{p^{L*}=0,\ p^{H*}} \geq \left[\frac{1 + \frac{\int_{\theta_L}^{\theta_{HL}} f(\theta)d\theta}{p^H f(\theta_{HL})}\left(\frac{\partial V^H(\theta_{HL},D)}{\partial \theta_{HL}} - \frac{\partial V^L(\theta_{HL},D)}{\partial \theta_{HL}}\right)}{f(\theta_L)\frac{\partial \theta_L}{\partial p^L}}\right]|_{p^{L*}=0,\ p^{H*}}$$

Corollary 2 explicitly states the necessary condition for freemium to be optimal for the running example.

Corollary 2. *With $\alpha^H \neq \alpha^L$, $V^i(\theta, D) = \theta q^i + \alpha^i D$ for $i \in \{L, H\}$, and $\theta \in U[0,1]$, freemium can be the optimal equilibrium strategy when the following necessary condition is satisfied:*

$$\alpha^H - \alpha^L \geq \left(q^L - \alpha^L\right)\left[1 + \frac{\theta^*_{HL} - \theta^*_L}{p^{H*}}\left(q^H - q^L\right)\right]$$

where $0 < \theta^*_L < \theta^*_{HL} < 1$ *and*

$$\theta^*_L = \frac{u_0 - \alpha^L}{q^L - \alpha^L},$$

$$\theta^*_{HL} = \frac{\left(q^H - q^L\right)q^L + \alpha^H\left(u_0 - q^L\right) - \alpha^L\left(q^H - 2q^L + u_0\right)}{2\left(q^H - q^L\right)\left(q^L - \alpha^L\right)},$$

$$p^{H*} = \frac{q^L\left(\alpha^H - q^L\right) - \left(\alpha^H - \alpha^L\right)u_0}{2\left(q^L - \alpha^L\right)} + \frac{q^H}{2}.$$

From Proposition 2 and Corollary 2, we can see that freemium can be optimal only if the network effect differential $\alpha^H - \alpha^L$ is higher than some positive value.

To solve the profit maximization problem, when two products are offered, first we can get the prices (p^H, p^L) expressed by marginal consumer types (θ_{HL}, θ_L) by Lemma 1. Substituting the price function back to the profit function Π_{HL}, the optimal θ^*_L and θ^*_{HL} are the values that maximize Π_{HL} under the constraint $0 \leq \theta_L < \theta_{HL} < 1$ and $p^L \geq 0$, and then the corresponding p^{L*} and p^{H*} can be obtained. When only the high-end product is offered, we can solve for θ^*_H and thus p^{H*} following similar logic. For the running example, the profit function is concave in θ_k, with $k \in \{H, L, HL\}$; the optimal solutions, thus, can be obtained straightforwardly. The following Corollary 3 gives the optimal product and pricing strategies and therefore describes a sufficient condition for freemium to be optimal.

Corollary 3. *With* $\alpha^H \neq \alpha^L$, $V^i(\theta, D) = \theta q^i + \alpha^i D$ *for* $i \in \{L, H\}$, *and* $\theta \in U[0,1]$, *the firm's equilibrium product and pricing strategy can be characterized as follows :*

(a) When $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^{L*}=0, p^{H*}} \leq 0$, $0 < \frac{u_0 - \alpha^L}{q^L - \alpha^L} < \frac{\left(q^H - q^L\right) q^L - \alpha^L \left(q^H - 2q^L + u_0\right) + \alpha^H \left(u_0 - q^L\right)}{2(q^H - q^L)(q^L - \alpha^L)} < 1$, *and* $\Pi^*_{HL} \geq \frac{\left(q^H - u_0\right)^2}{4(q^H - \alpha^H)}$, *the firm offers two products with*

$$
\begin{aligned}
p^{L*} &= 0, \\
p^{H*} &= \frac{q^L \left(\alpha^H - q^L\right) - \left(\alpha^H - \alpha^L\right) u_0}{2 \left(q^L - \alpha^L\right)} + \frac{q^H}{2}, \\
\theta^*_L &= \frac{u_0 - \alpha^L}{q^L - \alpha^L}, \\
\theta^*_{HL} &= \frac{\left(q^H - q^L\right) q^L - \alpha^L \left(q^H - 2q^L + u_0\right) + \alpha^H \left(u_0 - q^L\right)}{2 \left(q^H - q^L\right) \left(q^L - \alpha^L\right)}, \\
\Pi^*_{HL} &= \frac{\left[\alpha^L \left(q^H - u_0\right) - q^L \left(q^H - q^L\right) + \alpha^H \left(u_0 - q^L\right)\right]^2}{4 \left(q^H - q^L\right) \left(q^L - \alpha^L\right)^2}.
\end{aligned}
$$

(b) When $0 \leq \frac{p^{L*} + u_0 - \alpha^L}{q^L - \alpha^L} < \frac{p^{H*} - p^{L*} - \left(\alpha^H - \alpha^L\right) \frac{q^L - p^{L*} - u_0}{q^L - \alpha^L}}{q^H - q^L} < 1$, $0 < p^{L*} < p^{H*}$, *and* $\Pi^*_{HL} \geq \frac{\left(q^H - u_0\right)^2}{4(q^H - \alpha^H)}$, *the firm offers both products with positive prices and*

$$
\begin{aligned}
\theta^*_L &= \frac{\left(\alpha^H - \alpha^L\right) \left(\alpha^H - \alpha^L - q^H + q^L\right) + 2 \left(q^H - q^L\right) \left(\alpha^H + \alpha^L - u_0 - q^L\right)}{\left(\alpha^H - \alpha^L\right)^2 - 4 \left(q^H - q^L\right) \left(q^L - \alpha^L\right)} \\
\theta^*_{HL} &= \frac{\left(\alpha^H - \alpha^L\right)^2 + \alpha^L \left(u_0 + 2q^H - 3q^L\right) + \alpha^H \left(q^L - u_0\right) - 2q^L \left(q^H - q^L\right)}{\left(\alpha^H - \alpha^L\right)^2 - 4 \left(q^H - q^L\right) \left(q^L - \alpha^L\right)}, \\
\Pi^*_{HL} &= \frac{\left(q^H - q^L\right) \left[\alpha^L \left(q^H - c\right) - \alpha^H \left(q^L - c\right) + \left(2c - q^H\right) q^L - c^2\right]}{\left(\alpha^H - \alpha^L\right) + 4 \left(\alpha^L - q^L\right) \left(q^H - q^L\right)}.
\end{aligned}
$$

(c) Otherwise the firm offers only the high-end product with $p^{H*} = \frac{q^H - u_0}{2}$, $\theta^*_H = \frac{q^H + u_0 - 2\alpha^H}{2(q^H - \alpha^H)}$, $\Pi^*_H = \frac{\left(q^H - u_0\right)^2}{4(q^H - \alpha^H)}$.

Under asymmetric network effects (i.e., $\frac{\partial V^H(\theta_{HL}, D)}{\partial D} - \frac{\partial V^L(\theta_{HL}, D)}{\partial D}$ has to be greater than a certain positive threshold), the firm can indeed increase total profit by offering the low-end product for free. The key intuition is as follows. As the firm cuts its low-end price to expand demand, both high- and low-end products become more attractive. However, due to the difference in network effects $\frac{\partial V^H(\theta_{HL},D)}{\partial D}$ and $\frac{\partial V^L(\theta_{HL},D)}{\partial D}$, the high-end product becomes relatively more attractive. Put differently, the incentive compatibility constraint for the marginal consumer can be less tight when the network size is larger, due to asymmetric network effects. As such, when $\frac{\partial V^H(\theta_{HL},D)}{\partial D} - \frac{\partial V^L(\theta_{HL},D)}{\partial D}$ is large enough, holding p^H fixed, decreasing p^L may lead to higher demand for the high-end product and therefore higher profit. The premium from the paid users is indeed sufficient to pay for the losses.

We can also see that in the running example, with asymmetric network effects, $q^H - q^L$ needs to be small enough. Intuitively, if $q^H - q^L$ is too large, even with $p^L = 0$, the low-end product quality q^L is not high enough to attract many new customers (i.e., θ_L is too large), and is thus unable to create high enough network benefit. In this case, even a fairly large asymmetry between α^H and α^L cannot induce freemium as the optimal strategy (the necessary condition cannot be satisfied). This condition makes it sufficient for freemium to be adopted as an equilibrium strategy. Please see the appendix for a more detailed discussion.

Figure 1: Product line strategy under network effects ($q^H = 0.8$, $q^L = 0.2$, $u_0 = 0.06$)

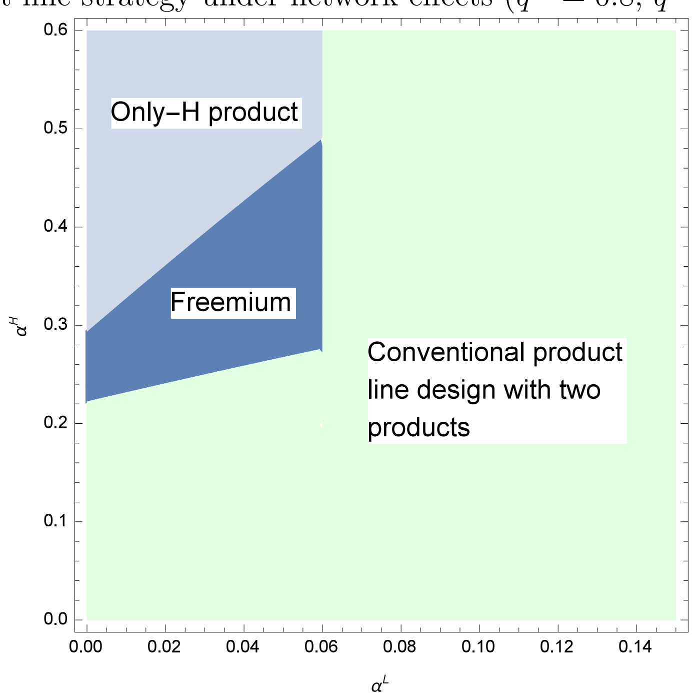


Figure 1 illustrates these equilibrium strategies in the parameter space for the running example. Simply put, when the difference between the network effects of low- and high-end products is relatively large, the firm has less pressure from cannibalization and focuses on network effects. This leads to the freemium strategy, wherein the firm forgoes the profit from the low-end customers and considers them a subsidy to the premium users. When the network effects of the low-end product becomes stronger, the firm pursues a conventional product line strategy, where prices for both products are positive. When the high-end product's network effects (i.e. α^H) are very high, the firm would be better off by offering only the high-end product.

It is worth noting that, as shown in Figure 1, if we decrease α^L while fixing α^H, the optimal strategy may move from freemium to offering only the high-end product. This may seem puzzling at first glance because a lower α^L leads to a larger asymmetry between the network effects, which further relaxes the cannibalization pressure. However, another effect of lowering α^L is that the firm would find it harder (costlier) to attract consumers to use the low-end product, and therefore, offering only the high-end product becomes more attractive. Mathematically, as shown by the necessary condition in Proposition 2 and Corollary 2, when α^L decreases Δ, the left hand side of the inequality increases Δ, but the right hand side increases more than Δ, making it harder to satisfy the inequality. Similarly, if we increase α^H while fixing α^L, the optimal strategy may also move from freemium to offering only the high-end product. This is because that the profit of offering only the high-end product increases in α^H. Therefore, even though the necessary conditon for freemium to be optimal is easier to satisfy with a larger α^H, the sufficient condition, which requires $\Pi^*_{HL}|_{p^{L*}=0,p^{H*}} > \Pi^*_H$, is harder to satisfy.

Table 1 provides a summary of our main insights. We describe the product line choices under each type of equilibrium and provide their existence conditions.

Table 1: Summary of product and pricing choices in equilibria

Type of equilibrium	*Features*	*Exists under no n.e.?*	*Exists under uniform n.e.?*	*Exists under asymmetric n.e.?*
(1) *Conventional product line*	*Anderson & Celik* 2014 $p^{H*} > p^{L*} > 0$	✓	✓	✓
(2) *Freemium*	$p^{L*} = 0$, $p^{H*} > 0$	×	×	✓
(3) *Only one product*	$p^{*} > 0$	✓	✓	✓

It should be noted, that the phenomenon of freemium can be considered as a special case of complementary good pricing. A monopoly who sells two complementary goods has incentives to lower each product's price if this leads to higher sales of the other product. If the complementarity by one product to the other product is stronger, the firm may lower the former product's price below marginal cost in order to profit from the sales of the latter product. Our model essentially builds on this insight in a vertical differentiation scenario. We argue that in a vertical product line, the cannibalization effect co-exists with the complementarity effect. In the case of network effects, the network effects have to be asymmetric so that the complementarity effect outweighs the cannibalization effect when the firm lowers the entry-level product's price. The ecomomic intuition, however, may indeed have broader applications.[8]

1.5 Endogenous quality decisions

The analyses so far speak to the conditions under which freemium is or is not optimal in a product line with given qualities and network effects. In this section, we consider the endogeneous determination of the quality levels and network effects. This is a technically challenging exercise, and we approach it by considering two alternative formulations. Section 1.5.1 considers the running example introduced in Section 1.3, with uniform distribution of

[8]We thank an evaluator for this insightful comment.

consumer type and linear utility function. We allow q^i to be a decision variable and let $C(q^i)$ denote the firm's marginal production cost of products with quality q^i. This formulation allows us to study the optimal determination of standalone qualities (q^i) when network effects are given. In Section 1.5.2, we fully endogenize both q^i as well as α^i in a discrete segment model, which corresponds to the widely used model of two-type screening.

Remarkably, the key insights in Section 1.4 hold true in all the alternative formulations. Thus, this section could also be considered as a robustness check, while we make more variables endogeneous in progressively simpler models. In both subsections, we relegate the proofs to the appendix and rely on numerical methods to generate the graphical illustrations.

1.5.1 Endogenous quality

With q^H, q^L, p^H, p^L all endogenized, we take a closer look at the equilibrium levels of q^i as a function of network effects. That is, how should the product qualities shift with network effects?

To proceed, we need the following two assumptions in addition to $AI \sim AIV$.

AV. $V^i(\theta, D)$ is concave in q^i, $\forall i \in \{H, L\}$.

AVI. $C(q^i)$ is convex in q^i, $\forall i \in \{H, L\}$.

With the above two assumptions together with $AI \sim AIV$, we find that the firm should respond to higher α^L by increasing q^L in conventional product line. However, it should respond to higher α^L by reducing quality provision q^L while adopting freemium. Proposition 3 illustrates the firm's choices of quality in the conventional product line and freemium regime, with the running example satisfying $AI \sim AIV$ and marginal cost satisfying *AVI.*

Proposition 3. *With valuation function $V^i(\theta, D) = \theta q^i + \alpha^i D$, marginal production cost $C(q^i) = c \cdot (q^i)^2$ and $\theta \sim U[0,1]$, when conventional product line is the optimal strategy, the firm offers both products with $q^{H*} = \frac{\theta^*_{HL}}{2c}$, $q^{L*} = \frac{\theta^*_L + \theta^*_{HL} - 1}{2c}$, and q^{L*} increases in α^L; when freemium is the optimal strategy, the firm offers both products with $q^{H*} = \frac{\theta^*_{HLf}}{2c}$, $q^{L*} = \frac{u_0 - \alpha^L\left(1 - \theta^*_{Lf}\right)}{\theta^*_{Lf}}$, and q^{L*} decreases in α^L.*

As explained previously, when a firm pursues the freemium strategy, it suffers a loss on the low-valuation segment and recuperates the lost profit from the high-valuation segment. Because the firm desires the low-valuation consumers for the sole purpose of enlarging its network size, it should always supply the least level of quality that is sufficient to induce purchase. As α^L increases, the minimal required quality level decreases accordingly, leading to lower quality provision. Broadly speaking, quality and network effects are substitutes for the low-end product. The firm always seeks the least costly way to attract the low-end customers, and as one dimension gets higher, it decreases its investment in the other dimension.

1.5.2 Endogenous quality and network effect

In the baseline model, we examined a general utility functional form as well as a general consumer distribution. In Section 1.5.1, the product quality is endogenized. Next, we endogenize both quality and network effect decisions. In reality, the network effect—as a product attribute—may also be the firm's endogenous decision. For example, in social games, the game designer (the firm) can endogenously decide how much network effect the paid users and free users can get, by designing the frequency of interaction between different types of players. In the case of Dropbox, the firm can make sharing more or less convenient so that different products deliver different network effect.

To keep the analysis tractable, we consider a discrete distribution of consumers on the demand side. Namely, there are two segments of consumers, with high and low valuation for the product quality as well as network benefit. Each consumer is characterized by a taste parameter $\theta \in \{\theta_H, \theta_L\}$, where $\theta_H > \theta_L$. A fraction λ of consumers belong to type θ_H, who have higher valuation for the firm's products. A fraction $1 - \lambda$ of consumers belong to type θ_L. There may be some debate on the formulation of the running example regarding whether the taste parameter (i.e., θ) affects only the valuation of the standalone quality, or the valuation of both quality and network benefit. In short, both formulations satisfy

$AI \sim AIV$, therefore, the results derived for the general model in Section 1.4 hold for both specifications. To demonstrate our results hold for both cases, we therefore offer a formulation in this extension different from that used in the previous running example. For a customer with taste parameter θ, her valuation from consuming product i is:

$$V^i(\theta, \alpha^i, D) = \theta(q^i + \alpha^i D),$$

where $D \leq 1$ is the total number of users who buy from the firm's product line, namely $D = \sum_{i \in \{H,L\}} D^i$. As such, we assume that each user derives network effects from all other users in the firm's network. The total magnitude of network effects depends on the network size and product design. The firm sets price $p^i > 0$ for each product i. To guarantee the existence of interior solutions, we assume that the marginal cost of serving a consumer is increasing in both q^i and α^i quadratically, i.e., $C\left(q^i, \alpha^i\right) = c\left(q^i\right)^2 + s\left(\alpha^i\right)^2$. The firm's product line profit is thus:

$$\Pi = \sum_{i \in \{H,L\}} D^i \left[p^i - c\left(q^i\right)^2 - s\left(\alpha^i\right)^2\right].$$

We can see that all of $AI \sim AVI$ are satisfied. In this discrete case, we also assume that the high-type consumers have positive valuation for the low-end product at price zero.[9] Proposition 4 spells out the optimal product line design under uniform network effect.

Proposition 4. *When $\alpha^H = \alpha^L = \alpha$, the equilibrium product-line strategy can be characterized as follows:*

$$\text{When } (I)\begin{cases} \frac{\theta_L^2 - \lambda^3\theta_H^2}{4s} - (1-\lambda)u_0 > 0 \\ \lambda\theta_H\theta_L > \theta_L^2 \geq 2su \end{cases} \quad \text{or } (II)\begin{cases} \frac{(\lambda\theta_H - \theta_L)^2}{4c(1-\lambda)} + \frac{2\theta_L - \lambda^3\theta_H^2}{4s} - (1-\lambda)u_0 > 0 \\ s\left[\theta_H\theta_L\lambda + 2cu_0(1-\lambda)\right] \leq \theta_L^2(c - c\lambda + s) \\ \lambda \leq \frac{\theta_L}{\theta_H} \end{cases} \quad \text{the}$$

[9]In appendix B, we explain specifically for the discrete case where this assumption does not hold, which means that the highest-type consumers will not find the low-end product worth trying even if it is offered at zero price. This is implausible if not impossible according to those successful freemium products offered in the market.

firm offers both products with price $p^{H*} > p^{L*} > 0$.[10] *The corresponding optimal qualities, network intensity and firm profit are:*

$$
\begin{aligned}
q^{L*} &= \begin{cases} 0 & ,when\ (I)\ holds \\ \frac{\theta_L - \theta_H \lambda}{2c(1-\lambda)} & ,when\ (II)\ holds \end{cases} \\
q^{H*} &= \frac{\theta_H}{2c},\ \alpha^* = \frac{\theta_L}{2s} \\
\Pi^*_{HL} &= \begin{cases} \frac{\lambda \theta_H^2}{4c} + \frac{\theta_L^2}{4s} - u_0 & ,when\ (I)\ holds \\ \frac{\theta_L^2 + \lambda \theta_H^2 - 2\lambda \theta_H \theta_L}{4c(1-\lambda)} + \frac{\theta_L^2}{4s} - u_0 & ,when\ (II)\ holds \end{cases}
\end{aligned}
$$

Otherwise, the firm offers only a high-end product. The price, quality, and equilibrium profit are:

$$
\begin{aligned}
p^* &= \frac{\theta_H^2}{2c} + \frac{\theta_H^2 \lambda^2}{2s} - u_0, \\
q^* &= \frac{\theta_H}{2c},\ \alpha^* = \frac{\lambda \theta_H}{2s}, \\
\Pi^*_H &= \lambda \left(\frac{\theta_H^2}{4c} + \frac{\lambda^2 \theta_H^2}{4s} - u_0 \right).
\end{aligned}
$$

The following Figure 2 shows regions for optimal product line strategies under zero network effects as well as uniform network effects. The figure 2(a) corresponds to the case where $\alpha = 0$, i.e., consumers derive no network benefit from using the product. The figure 2(b) depicts the regions for optimal strategies for the case with $\alpha > 0$, namely, the network effects are positive and symmetric.

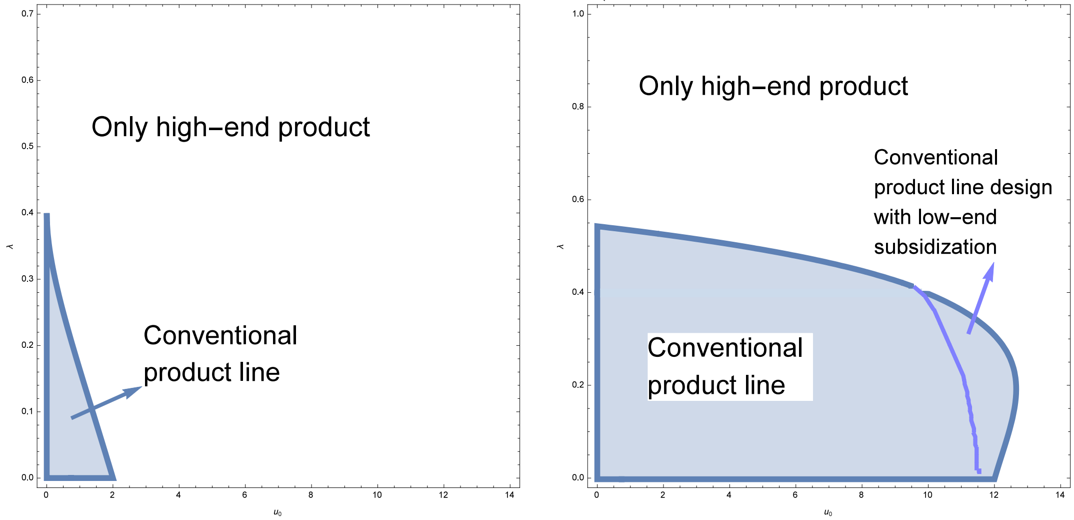
From the above figures, we can see that freemium is never an optimal strategy when network effects are zero or symmetric, consistent with the results of our baseline model.

Next, we consider the case in which the firm can design its high-end and low-end products to deliver different levels of network effects. Proposition 5 states the optimal product line design when the network effects can be designed as asymmetric.

Proposition 5. *When the firm can set α^H and α^L at different levels, the equilibrium product-*

[10] The exact expressions of optimal prices are provided in the appendix.

Figure 2: Regions for optimal strategies ($c = 0.5, s = 0.1, \theta_H = 5, \theta_L = 2$)



(a) Case without network effect (b) Case with uniform network effect

line strategy can be characterized as follows:

When $\frac{\theta_H^2}{4s}(\lambda - \lambda^3) + \lambda u_0(1 - \frac{\theta_H}{\theta_L}) - \frac{csu_0^2(1-\lambda)}{\theta_L^2(c+s)} \geq 0$ *and* $\frac{2csu_0(1-\lambda)}{\theta_L{}^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 > 0$*, the firm adopts freemium strategy with:*

$$p^{L*} = 0,\ p^{H*} = \frac{\theta_H^2}{2}\left(\frac{1}{s} + \frac{1}{c}\right) - \frac{u_0\theta_H}{\theta_L}$$

The corresponding optimal qualities, network effects and profit are:

$$q^{L*} = \frac{su_0}{\theta_L\,(c+s)},\ q^{H*} = \frac{\theta_H}{2c}$$
$$\alpha^{L*} = \frac{cu_0}{\theta_L\,(c+s)},\ \alpha^{H*} = \frac{\theta_H}{2s}$$
$$\Pi_F^* = \frac{\theta_H^2\lambda}{4}\left(\frac{1}{s} + \frac{1}{c}\right) - \frac{cs\,(1-\lambda)\,u_0^2}{\theta_L^2\,(c+s)} - \frac{\lambda u_0\theta_H}{\theta_L}$$

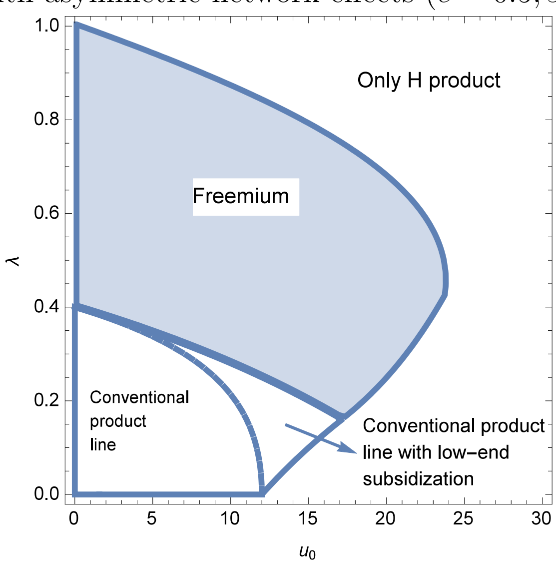
When $\frac{2csu_0(1-\lambda)}{\theta_L{}^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 \leq 0$ *and* $\frac{(\theta_L-\theta_H\lambda)^2}{4c(1-\lambda)}\left(\frac{1}{s} + \frac{1}{c}\right) + \frac{\theta_H^2}{4s}\left(\lambda - \lambda^3\right) - u_0\left(1-\lambda\right) > 0$*, the firm launches two products with* $p^{H*} > p^{L*} > 0$*. The corresponding qualities, network effects* *and profit are:*

$$q^{L*} = \frac{\theta_L - \theta_H\lambda}{2c(1-\lambda)},\ q^{H*} = \frac{\theta_H}{2c},$$
$$\alpha^{L*} = \frac{\theta_L - \theta_H\lambda}{2s(1-\lambda)},\ \alpha^{H*} = \frac{\theta_H}{2s},$$
$$\Pi^*_{HL} = \frac{\theta_L^2 + \theta_H^2\lambda - 2\theta_L\theta_H\lambda}{4c(1-\lambda)} + \frac{\theta_H^2\lambda}{4s} + \frac{(\theta_L - \theta_H\lambda)^2}{4s\,(1-\lambda)} - u_0.$$

Otherwise, the firm provides only the high-end product, with price, quality, network effect and profit as:

$$p^* = \frac{\theta_H^2}{2c} + \frac{\lambda^2\theta_H^2}{2s} - u_0,$$
$$q^* = \frac{\theta_H}{2c},\ \alpha^* = \frac{\lambda\theta_H}{2s},$$
$$\Pi^*_H = \lambda\left(\frac{\theta_H^2}{4c} + \frac{\lambda^2\theta_H^2}{4s} - u_0\right).$$

Figure 3: Strategies with asymmetric network effects ($c = 0.5, s = 0.1, \theta_H = 5, \theta_L = 2$)



From Figure 3, we can see that under asymmetric network effects, the firm can indeed increase product-line profit by offering the low-end product for free. However, it should be noted that even when asymmetric network effects are present, freemium is not always an optimal strategy. When both λ and u_0 are relatively low, the firm segments the market via a conventional product line. In this case, p^L once again corresponds to the efficient price.

Given the above results, we compare the optimal quality decision under freemium and conventional product line design. We use the term "efficient quality", denoted by q^{i0*} with $i \in \{H, L\}$, to refer to the quality level that maximizes single product profit (therefore social welfare) under the complete information benchmark.

Corollary 4. *Across all equilibria, the quality of the high-end product is always set at the efficient level, i.e.* $q^{H*} = q^{H0*}$*;*

When both segments are served with positive prices, the quality of the low-end product is always below the efficient level, i.e. $q^{L*} < q^{L0*}$*;*

When the freemium strategy is adopted, the quality of the low-end product can be lower, equal to, or even greater than the efficient level, i.e., $q^{L*} = \frac{su_0}{\theta_L(c+s)} > q^{L0*}$ *can hold;*

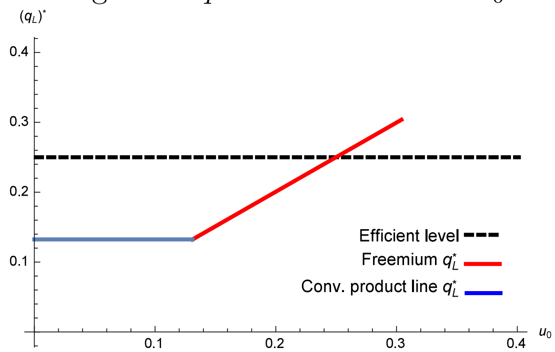
The quality difference $\Delta q = q^{H*} - q^{L*}$ *is smaller under the freemium strategy than that under conventional product line strategy with uniform network effects.*[11]

For freemium to be optimal, previous studies have recognized the importance of offering a balanced set of features in a firm's free product (Lee et al., 2013; Kumar, 2014). Kumar (2014) stated that for freemium to work, the free offering has to be "compelling enough" to attract new users, but it cannot be "too rich" such that people stick to the free product. This insight is confirmed by our analysis. Whenever a freemium strategy is pursued, the firm has to strike a balance between getting consumers on board and minimizing cannibalization.

At the same time, this insight alone is not enough for the design of an optimal product line. Corollary 4 speaks to the importance of distinguishing between features that contribute to a product's standalone quality and features that contribute to its network effects. It states that when a firm pursues the freemium strategy, it should choose a standalone quality level for its low-end product that is higher than what it would choose in a conventional product line. In other words, a firm pursuing the freemium strategy should offer more features at

[11] Here we compare the quality difference, assuming that the relevant parameter values are the same.

Figure 4: q^{L*} as a function of u_0



a lower price (i.e., zero)! This reduces the quality differentiation within the product line. To compensate, the firm should design the products such that they offer different levels of network effects. In other words, it is not just the number of free features that determines freemium's viability, but also which features are included in the free version. The prescription of higher standalone quality may seem counterintuitive at first, but a firm should realize that it is precisely a high standalone quality that allows the firm to choose lower network effects for the low-end product. Lower network effects are the key to minimizing cannibalization.

Figure 4 provides a simple illustration of this idea. It plots the equilibrium level of q^{L*} against the value of u_0. As u_0 increases, the equilibrium shifts from a conventional product line to the freemium equilibrium, and q^{L*} increases accordingly.

1.6 Concluding Remarks

This paper studies a monopolist's product and pricing strategy under network effects. We are particularly interested in the optimality of the freemium strategy. We seek to answer two questions. First, what are the necessary conditions for the optimality of freemium? Our results point to the asymmetry in network effects as the determining factor. Second, what are the principles that should guide the design of freemium? Our results add to the previous literature and speak to the distinction between a product's "baseline" quality and the network effects its users receive. Compared to a conventional product line, in a freemium equilibrium, the firm should offer relatively higher baseline quality but low network effects on its low-end product.

Of course, as we have reviewed in Section 1.2, our analysis provides only one of many explanations of the optimality of freemium. Kamada and Ory (2015), for example, focused on the alternative explanation where freemium motivates word-of-mouth during the diffusion process. Other papers we reviewed in Section 1.2 considered the importance of consumer learning and free trial. In addition, competition may also play a key role in driving the price of the low-end product to the marginal cost. These explanations are clearly not mutually exclusive. Which explanation serves as the most likely explanation to the observed freemium depends on the context that is being considered. Our framework applies best to a scenario where the product is beyond the diffusion stage, and competition is not intense. When competition is relevant, the same mechanism may still be at work, but competition itself creates a downward pressure on the prices and provides the focal firm strong incentives to pull out the low-end product altogether. The exact effect is an interesting topic for future research.

Freemium has become an immensely popular business model among start-ups, especially in the Internet sector. Without doubt, providing a product for free is an effective way to expand a firm's user base. As many entrepreneurs have rightly believed, expanding a firm's user base is of ultimate importance for industries with network effects. However, our analysis points out that a firm should exercise caution when it is tempted to jump on the freemium bandwagon in order to "get users on board." For freemium to work, a firm has to understand the subtleties of network effects in its market. When network effects are not strong enough or are uniform across segments, offering an entry level product for free does no good to the product line profit. Freemium expands a firm's market share but severely limits its ability to create enough margin from its paid users. This scenario may sound familiar to many firms who are frustrated by the disappointing number of paying users under their freemium strategy. Instead, these firms should heed the wisdom of Mussa and Rosen (1978), and segment the market via a conventional product line. A conventional product line consists of a low quality product that is sold at an efficient price, instead of an entry level product sold at zero price.

Our results in the extensions also provide guidelines to firms who should certainly adopt the freemium strategy. As previous studies have pointed out, a firm should provide an intermediate number of features in its free product. Moreover, it is not just the number of features that matters, but also which features the firm should provide. Our analysis prescribes that a firm should in fact be quite generous with features that enhance a product's "baseline quality" – that is, the value of a product when it is used alone. However, the firm should deliberately limit the features that bring network benefits to users. In a well designed freemium menu, "paying for upgrade" should in fact be "paying for network effects."

Our study focuses on a scenario wherein product line profit is the main driver of firm strategy. In doing so, we leave out many behavior factors that are nonetheless relevant to the freemium strategy. For example, offering a product for free can induce greater word-of-mouth and speed up its adoption (e.g., Kamada and Ory 2015). It would be interesting to combine the two perspectives and investigate the dynamics of the freemium strategy. This exercise may lead to a "taxonomy" of the freemium strategy and elucidate the possible motivations behind offering a free product. Second, it is interesting to extend the current model and exam the possibilities of advertising income. Advertising income should give the firm stronger incentives to provide the free product, and a firm should strike a balance between lower fees and higher ad revenue. Third, competition is a relevant factor in many high tech markets. Even though product-line competition brings considerable complexities to the model, it remains a meaningful direction for future research. Finally, it is of some interest to generalize the model into a case of user subsidization, where negative price is possible. When the firm is able to subsidize the users, the price for the entry-level product is not constrained to be non-negative. Although it is relatively easy to persuade adoption with subsidies, it is much harder to induce actual usage. It is of managerial interests to explore strategies that the firm could follow when network effects stem mostly from actual usage.

References

[1] Anderson, S. and Celik, L. (2014). Product line design. *Working Paper*.

[2] Banerji, A. and Dutta, B. (2009). Local network externalities and market segmentation. *International Journal of Industrial Organization*, 27(5):605–614.

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

[4] Church, J. and Gandal, N. (1992). Network effects, software provision, and standardization. *The Journal of Industrial Economics*, 40(1):85–103.

[5] Desai, P., Kekre, S., Radhakrishnan, S., and Srinivasan, K. (2001). Product differentiation and commonality in design: balancing revenue and cost drivers. *Management Science*, 47(1):37–51.

[6] Desai, P. S. (2001). Quality segmentation in spatial markets: when does cannibalization affect product line design? *Management Science*, 20(3):265–283.

[7] Dey, D., Lahiri, A., and Liu, D. (2013). Consumer Learning and time-locked trials of software products. *Journal of Management Information Systems*, 30(2):239–268.

[8] Farrell, J. and Klemperer, P. (2007). Coordination and lock-in: competition with switching costs and network effects. *Handbook of Industrial Organization*, 3(Chapter 31):1970–1994.

[9] Fjeldstad, Ø., Riis, C., and Moen, E. R. (2010). Competition with local network externalities. *Working Paper*.

[10] Hagiu, A. (2006). Pricing and commitment by two-sided platforms. *The RAND Journal of Economics*, 37(3):720–737.

[11] Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(0108):44–74.

[12] Jing, B. (2007). Network externalities and market segmentation in a monopoly. *Economics Letters*, 95(1):7–13.

[13] Jing, B. and Zhang, Z. J. (2011). Product line competition and price promotions. *Quantitative Marketing and Economics*, 9(3):275–299.

[14] Jones, R. (2015). Freemium debate heats up. *Music Week*.

[15] Kamada, Y. and Ory, A. (2015). Encouraging word of mouth : Free contracts , referral programs , or both ? *Working Paper*.

[16] Katz, M. and Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75:424–440.

[17] Katz, M. L. and Shapiro, C. (1986). Technology adoption in the presence of network externalities. *Journal of Political Economy*, 94:822–841.

[18] Kumar, V. (2014). Making "freemium" work. *Harvard Business Review*, (May):701–703.

[19] Lee, C., Kumar, V., and Gupta, S. (2013). Designing freemium: A model of consumer usage, upgrade, and referral dynamics. *Working Paper*.

[20] Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *The RAND Journal of Economics*, 15(2):171–196.

[21] Moorthy, K. (1988). Product and price competition in a duopoly. *Marketing Science*, 7(2):141–168.

[22] Moorthy, K. and Png, I. (1992). Market segmentation, cannibalization, and the timing of product introductions. *Management Science*, 38(3):345–359.

[23] Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18:301–317.

[24] Needleman, S. and Loten, A. (2012). When freemium fails. *The Wall Street Journal*.

[25] Niculescu, M. and Wu, D. (2014). Economics of free under perpetual licensing: Implications for the software industry. *Information Systems Research*, 25(1):173–199.

[26] Orhun, A. (2009). Optimal product line design when consumers exhibit choice set-dependent preferences. *Marketing Science*, 28(5):868–886.

[27] Riggins, F. (2003). Market segmentation and information development costs in a two-tiered fee-based and sponsorship-based web site. *Journal of Management Information Systems*, 19(3):69–86.

[28] Rochet, J. and Tirole, J. (2006). Two-sided markets: a progress report. *The RAND Journal of Economics*, 37(3):645–667.

[29] Shampanier, K., Mazar, N., and Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing science*, 26(6):742–757.

[30] Varian, H. R. and Shapiro, C. (1998). Versioning: The smart way to sell information. *Harvard Business Review*, 107(6):107.

[31] Wagner, T. M., Benlian, A., and Hess, T. (2014). Converting freemium customers from free to premium–the role of the perceived premium fit in the case of music as a service. *Electronic Markets*, 24(4):259–268.

[32] Wei, X. D. and Nault, B. R. (2013). Experience information goods: "Version-to-upgrade". *Decision Support Systems*, 56:494–501.

[33] Zetlin, M. (2013). Rhapsody CEO: Freemium doesn't work. *www.inc.com*.

[34] Zhang, K. and Sarvary, M. (2015). Differentiation with user-generated content. *Management Science*, 61(4):898–914.

Appendix A: Proofs.

Proof. ***Footnote 5.***

Below we prove $\forall D$, $\frac{\partial\left[V^H(\theta,D)-V^L(\theta,D)\right]}{\partial\theta} > 0$ given the following two conditions:

$$\forall\theta, \frac{\partial\left[V^H(\theta,0)-V^L(\theta,0)\right]}{\partial\theta} > 0(*) \text{ and } \forall\theta, D, \frac{\partial\left[\frac{\partial V^H(\theta,D)}{\partial D}-\frac{\partial V^L(\theta,D)}{\partial D}\right]}{\partial\theta} \geq 0(**).$$

We can write

$$\begin{aligned} V^H(\theta,D)-V^L(\theta,D) &= \left[V^H(\theta,0)+\int_0^D \frac{\partial V^H(\theta,t)}{\partial t}dt\right]-\left[V^L(\theta,0)+\int_0^D \frac{\partial V^L(\theta,t)}{\partial t}dt\right] \\ &= V^H(\theta,0)-V^L(\theta,0)+\int_0^D \frac{\partial\left[V^H(\theta,t)-V^L(\theta,t)\right]}{\partial t}dt \end{aligned}$$

then given $(*)$ and $(**)$ we have

$$\begin{aligned} \frac{\partial\left[V^H(\theta,D)-V^L(\theta,D)\right]}{\partial\theta} &= \frac{\partial\left[V^H(\theta,0)-V^L(\theta,0)+\int_0^D \frac{\partial\left[V^H(\theta,t)-V^L(\theta,t)\right]}{\partial t}dt\right]}{\partial\theta} \\ &= \frac{\partial\left[V^H(\theta,0)-V^L(\theta,0)\right]}{\partial\theta}+\int_0^D \frac{\partial\left[\frac{\partial V^H(\theta,D)}{\partial D}-\frac{\partial V^L(\theta,D)}{\partial D}\right]}{\partial\theta}dt \\ &> 0 \end{aligned}$$

□

Proof. ***Lemma 1.***

We used proof by contradiction for Lemma 1. Please refer to the main text for the complete proof. □

Proof. ***Proposition 1.***

To prove Proposition 1, we show that when $\frac{\partial V^H(\theta,D)}{\partial D} = \frac{\partial V^L(\theta,D)}{\partial D} \geq 0$, the necessary condition for freemium to be optimal (i.e., $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^L=0} \leq 0$) is violated.

Let us now derive the general expression of Π as a piecewise function of p^H and p^L. Consider θ_L,θ_H and θ_{HL} which are determined by

$$V^L(\theta_L, D) - p^L = u_0$$
$$V^H(\theta_{HL}, D) - p^H = V^L(\theta_{HL}, D) - p^L$$
$$V^H(\theta_H, D) - p^H = u_0$$

Clearly, θ_L,θ_H and θ_{HL} are implicit functions of p^H and p^L. As a consequence, the profit function Π is a function of p^H and p^L:

$$\Pi = \begin{cases} p^H & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \le 0 \;\&\; \theta_{HL} \le 0. \\ & \text{In this case } D^L = 0, D^H = 1. \\ p^L & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \le 0 \;\&\; \theta_{HL} \ge 1. \\ & \text{In this case } D^L = 1, D^H = 0. \\ 0 & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \ge 1 \;\&\; \theta_H \ge 1. \\ & \text{In this case } D^L = D^H = 0. \\ p^H D^H & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \ge 1 \;\&\; \theta_H \in (0,1)\,,\ \textit{or } 0 < \theta_H < \theta_L \le 1. \\ & \text{In this case } \ D^L = 0, D^H = \int_{\theta_H}^1 f(\theta)\, d\theta. \\ p^L D^L & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \in (0,1) \;\&\; \theta_{HL} \ge 1. \\ & \text{In this case } D^L = \int_{\theta_L}^1 f(\theta)\, d\theta, D^H = 0. \\ p^L D^L + p^H D^H & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \le 0 \;\&\; \theta_{HL} \in (0,1)\,,\ \textit{or } 0 < \theta_L < \theta_{HL} \le 1. \\ & \text{In this case } D^L = \int_{\theta_L \cdot I(\theta_L \ge 0)}^{\theta_{HL}} f(\theta)\, d\theta, D^H = \int_{\theta_{HL}}^1 f(\theta)\, d\theta. \end{cases}$$

$$= \begin{cases} p^H & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \le 0 \;\&\; \theta_{HL} \le 0. \\ p^L & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \le 0 \;\&\; \theta_{HL} \ge 1. \\ 0 & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \ge 1 \;\&\; \theta_H \ge 1. \\ p^H \int_{\theta_H}^1 f(\theta)\, d\theta & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \ge 1 \;\&\; \theta_H \in (0,1)\,, \\ & \ \textit{or } 0 < \theta_H < \theta_L \le 1. \\ p^L \int_{\theta_L}^1 f(\theta)\, d\theta & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \in (0,1) \;\&\; \theta_{HL} \ge 1. \\ p^L \int_{\theta_L \cdot I(\theta_L \ge 0)}^{\theta_{HL}} f(\theta)\, d\theta + p^H \int_{\theta_{HL}}^1 f(\theta)\, d\theta & \text{, when } p^H \text{and } p^L \text{are s.t. } \theta_L \le 0 \;\&\; \theta_{HL} \in (0,1)\,, \\ & \ \textit{or } 0 < \theta_L < \theta_{HL} \le 1. \end{cases}$$

First notice that, in equilibrium, θ_L^* cannot be smaller than 0, because the firm can always increase profit by increasing both p^H and p^L to make the type $\theta = 0$ indifferent between purchasing and not purchasing. Thus, we need to prove $\Pi_{HL} = p^L \int_{\theta_L}^{\theta_{HL}} f(\theta)\, d\theta + p^H \int_{\theta_{HL}}^1 f(\theta)\, d\theta$ violates the necessary condition $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^L=0} \le 0$, for any p^H as long as the demand schedule is $0 \le \theta_L < \theta_{HL} \le 1$.

Taking partial derivative of Π_{HL} w.r.t. p^L, we have

$$\frac{\partial \Pi_{HL}}{\partial p^L} = \int\limits_{\theta_L}^{\theta_{HL}} f(\theta)\, d\theta + p^L \left[f(\theta_{HL}) - f(\theta_L) \right] \left(\frac{\partial \theta_{HL}}{\partial p^L} - \frac{\partial \theta_L}{\partial p^L} \right) - p^H f(\theta_{HL}) \frac{\partial \theta_{HL}}{\partial p^L}$$

The sign of $\frac{\partial \Pi_{HL}}{\partial p^L}$ depend on the sign of $\frac{\partial \theta_L}{\partial p^L}$ and $\frac{\partial \theta_{HL}}{\partial p^L}$. Next we determine the signs of $\frac{\partial \theta_L}{\partial p^L}$ and $\frac{\partial \theta_{HL}}{\partial p^L}$.

Recall from equation (1), θ_L is implicitly given by $V^L(\theta_L, D) = p^L + u_0$. From assumption AII and $AIII$, we have $\frac{\partial V^i(\theta,D)}{\partial \theta} > 0$, $\frac{\partial V^i(\theta,D)}{\partial D} > 0$, and we also have D decreasing in p^L. Therefore, when p^L increases, we must have a higher θ_L to maintain the equality $V^L(\theta_L, D) = p^L + u_0$. In other words, we can get

$$\frac{\partial \theta_L}{\partial p^L} > 0 \tag{4}$$

Recall from equation (2), θ_{HL} is implicitly given by $V^H(\theta_{HL}, D) - V^L(\theta_{HL}, D) = p^H - p^L$. When $\frac{\partial V^i(\theta,D)}{\partial D} = 0$ or $\frac{\partial V^H(\theta,D)}{\partial D} = \frac{\partial V^L(\theta,D)}{\partial D}$, we have $V^H(\theta_{HL}, D) - V^L(\theta_{HL}, D) = V^H(\theta_{HL}, 0) - V^L(\theta_{HL}, 0)$. Namely, the demand change does not affect the valuation differential. So we have $V^H(\theta_{HL}, 0) - V^L(\theta_{HL}, 0) = p^H - p^L$. From assumption AIV, we have $\frac{\partial \left[V^H(\theta,0) - V^L(\theta,0) \right]}{\partial \theta} > 0$. Therefore when p^L decreases, we must have a higher θ_{HL} to maintain the equality $V^H(\theta_{HL}, 0) - V^L(\theta_{HL}, 0) = p^H - p^L$. In other words, we can get

$$\frac{\partial \theta_{HL}}{\partial p^L} < 0 \tag{5}$$

Because $\theta_{HL} > \theta_L$ and $f(\theta_{HL}) \geq 0$, with (5) we have

$$\begin{aligned} \frac{\partial \Pi_{HL}}{\partial p^L}\Big|_{p^L=0} &= \int\limits_{\theta_L}^{\theta_{HL}} f(\theta)\, d\theta - p^H f(\theta_{HL}) \frac{\partial \theta_{HL}}{\partial p^L}\Big|_{p^L=0} \\ &> 0 \end{aligned}$$

$\square$

Proof. ***Corollary 1.***

If the firm offers both products, according to equations (1) and (2), we have

$$\theta_L = \frac{p^L + u_0 - \alpha}{q^L - \alpha}$$
$$\theta_{HL} = \frac{p^H - p^L}{q^H - q^L}$$

With $\Pi_{HL} = p^L(\theta_{HL} - \theta_L) + p^H(1 - \theta_{HL})$, we have

$$\frac{\partial \Pi_{HL}}{\partial p^L} = \frac{2\left(p^H - p^L\right)}{q^H - q^L} - \frac{2p^L + u_0 - \alpha}{q^L - \alpha}$$
$$\frac{\partial \Pi_{HL}}{\partial p^H} = 1 - 2\frac{p^H - p^L}{q^H - q^L}$$

Taking first order conditions, we get

$$\theta^*_{HL} = \frac{1}{2},\ \theta^*_L = \frac{q^L + u_0 - 2\alpha}{2\left(q^L - \alpha\right)}$$
$$p^{H*} = \frac{q^H - u_0}{2},\ p^{L*} = \frac{q^L - u_0}{2}$$

As long as $0 \leq \theta^*_L < \theta^*_{HL}$, that is, $\begin{cases} q^L > \alpha > u_0 \\ q^L + u_0 \geq 2\alpha \end{cases}$, the above is the firm's optimal product and pricing strategy, that is, to offer both types of products, with $p^{H*} = \frac{q^H - u_0}{2}$, $p^{L*} = \frac{q^L - u_0}{2}$.

Otherwise, the firm offers only the high-end product, with $p^{H*} = \frac{q^H - u_0}{2}$, $\theta^*_H = \frac{u_0 + q^H - 2\alpha}{2(q^H - \alpha)}$ and $\Pi^*_H = \frac{\left(q^H - u_0\right)^2}{4(q^H - \alpha)}$.[12] □

[12] Note our implicit assumption is that the firm has incentive to enter the market, thus $\Pi^*_H > 0$, i.e., $q^H > \alpha$.

Proof. ***Proposition 2.***

The necessary condition for freemium to be optimal is $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^{L*}=0,p^{H*}} \leq 0$. We prove this can be satisfied when

$$\left[\frac{\partial V^H(\theta_{HL},D)}{\partial D} - \frac{\partial V^L(\theta_{HL},D)}{\partial D}\right]|_{p^{L*}=0,p^{H*}} \geq \left[\frac{1+\frac{\int_{\theta_L}^{\theta_{HL}} f(\theta)d\theta}{p^H f(\theta_{HL})}\left(\frac{\partial V^H(\theta_{HL},D)}{\partial\theta_{HL}} - \frac{\partial V^L(\theta_{HL},D)}{\partial\theta_{HL}}\right)}{f(\theta_L)\frac{\partial\theta_L}{\partial p^L}}\right]|_{p^{L*}=0,p^{H*}},$$

where the right-hand side is a positive value.

Recall from the proof of Proposition 1, we have

$$\frac{\partial \Pi_{HL}}{\partial p^L} = \int_{\theta_L}^{\theta_{HL}} f(\theta)\, d\theta + p^L \left[f(\theta_{HL}) - f(\theta_L)\right]\left(\frac{\partial \theta_{HL}}{\partial p^L} - \frac{\partial \theta_L}{\partial p^L}\right) - p^H f(\theta_{HL}) \frac{\partial \theta_{HL}}{\partial p^L}$$

According to the above equation, the necessary condition $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^{L*}=0,p^{H*}} \leq 0$ can be satisfied when

$$\frac{\partial \theta_{HL}}{\partial p^L}|_{p^{L*}=0,p^{H*}} \geq \frac{\int_{\theta_L}^{\theta_{HL}} f(\theta)\, d\theta}{p^H f(\theta_{HL})}|_{p^{L*}=0,p^{H*}} \tag{6}$$

Next we determine the value of $\frac{\partial \theta_{HL}}{\partial p^L}|_{p^{L*}=0,p^{H*}}$.

Recall from equation (2), we have $V^H(\theta_{HL}, D) - p^H = V^L(\theta_{HL}, D) - p^L$. By implicit function theorem, taking first derivative w.r.t. p^L on both sides, we can get

$$\frac{\partial V^H(\theta_{HL}, D)}{\partial \theta_{HL}} \frac{\partial \theta_{HL}}{\partial p^L} + \frac{\partial V^H(\theta_{HL}, D)}{\partial D} \frac{\partial D}{\partial p^L} = \frac{\partial V^L(\theta_{HL}, D)}{\partial \theta_{HL}} \frac{\partial \theta_{HL}}{\partial p^L} + \frac{\partial V^L(\theta_{HL}, D)}{\partial D} \frac{\partial D}{\partial p^L} - 1$$

Rearranging, we have

$$\frac{\partial \theta_{HL}}{\partial p^L} = \frac{\left[\frac{\partial V^H(\theta_{HL},D)}{\partial D} - \frac{\partial V^L(\theta_{HL},D)}{\partial D}\right]\left(-\frac{\partial D}{\partial p^L}\right) - 1}{\frac{\partial V^H(\theta_{HL},D)}{\partial \theta_{HL}} - \frac{\partial V^L(\theta_{HL},D)}{\partial \theta_{HL}}} \tag{7}$$

Substituting $\frac{\partial \theta_{HL}}{\partial p^L}$ expressed by equation (7) into the inequality (6), we have

$$\left[\frac{\left[\frac{\partial V^H(\theta_{HL},D)}{\partial D}-\frac{\partial V^L(\theta_{HL},D)}{\partial D}\right]\left(-\frac{\partial D}{\partial p^L}\right)-1}{\frac{\partial V^H(\theta_{HL},D)}{\partial \theta_{HL}}-\frac{\partial V^L(\theta_{HL},D)}{\partial \theta_{HL}}}\right]|_{p^{L*}=0,p^{H*}} \geq \frac{\int_{\theta_L}^{\theta_{HL}} f(\theta)\,d\theta}{p^H f(\theta_{HL})}|_{p^{L*}=0,p^{H*}}$$

Rearranging, we have

$$\left[\frac{\partial V^H(\theta_{HL},D)}{\partial D}-\frac{\partial V^L(\theta_{HL},D)}{\partial D}\right]|_{p^{L*}=0,p^{H*}} \geq \left[\frac{1+\frac{\int_{\theta_L}^{\theta_{HL}} f(\theta)d\theta}{p^H f(\theta_{HL})}\left(\frac{\partial V^H(\theta_{HL},D)}{\partial \theta_{HL}}-\frac{\partial V^L(\theta_{HL},D)}{\partial \theta_{HL}}\right)}{f(\theta_L)\frac{\partial \theta_L}{\partial p^L}}\right]|_{p^{L*}=0,p^{H*}}$$

Notice that the above rearrangement can be got because $\frac{\partial\left[V^H(\theta_{HL},D)-V^L(\theta_{HL},D)\right]}{\partial \theta_{HL}} > 0$ and $\frac{\partial D}{\partial p^L}|_{p^L=0} < 0$. The former is given by assumption AIV. Here we show the latter. Following the same logic in the proof of Proposition 1, when $\frac{\partial V^L(\theta,D)}{\partial D} \neq \frac{\partial V^H(\theta,D)}{\partial D}$, we still have $\frac{\partial \theta_L}{\partial p^L} > 0$. For Π_{HL}, we have $D=\int_{\theta_L}^1 f(\theta)\,d\theta$, and $\frac{\partial D}{\partial p^L}|_{p^L=0} = -f(\theta_L)\frac{\partial \theta_L}{\partial p^L} < 0$. With $\frac{\partial \theta_L}{\partial p^L} > 0$, $f(\theta) > 0$ and $\frac{\partial\left[V^H(\theta_{HL},D)-V^L(\theta_{HL},D)\right]}{\partial \theta_{HL}} > 0$ (AIV), it follows that $\left[\frac{1+\frac{\int_{\theta_L}^{\theta_{HL}} f(\theta)d\theta}{p^H f(\theta_{HL})}\left(\frac{\partial V^H(\theta_{HL},D)}{\partial \theta_{HL}}-\frac{\partial V^L(\theta_{HL},D)}{\partial \theta_{HL}}\right)}{f(\theta_L)\frac{\partial \theta_L}{\partial p^L}}\right]|_{p^{L*}=0,p^{H*}}$ is a positive value. □

Proof. ***Corollary 2 and Corollary 3.***

According to Lemma 1, we have

$$p^L = \theta_L\left(q^L-\alpha^L\right)+\alpha^L-u_0,$$
$$p^H = \left(q^H-q^L\right)\theta_{HL}+\alpha^H+\theta_L\left(q^L-\alpha^H\right)-u_0.$$

The total profit is

$$
\begin{aligned}
\Pi_{HL}\left(\theta_{HL},\theta_{L}\right) &= p^{H}\left(1-\theta_{HL}\right)+p^{L}\left(\theta_{HL}-\theta_{L}\right) \\
&= \left[\left(q^{H}-q^{L}\right)\theta_{HL}+\alpha^{H}+\theta_{L}\left(q^{L}-\alpha^{H}\right)-u_{0}\right]\left(1-\theta_{HL}\right) \\
&+ \left[\theta_{L}\left(q^{L}-\alpha^{L}\right)+\alpha^{L}-u_{0}\right]\left(\theta_{HL}-\theta_{L}\right) \\
s.t.\ & 0\leq\theta_{L}<\theta_{HL}<1 \\
& \theta_{L}\left(q^{L}-\alpha^{L}\right)+\alpha^{L}-u_{0}\geq 0
\end{aligned}
$$

We can see that $\Pi_{HL}\left(\theta_{HL},\theta_{L}\right)$ is concave in both θ_{HL} and θ_{L}. Therefore, we can get the global optimal $\theta_{HL}^{*},\theta_{L}^{*}$ that maximize Π_{HL}, by employing the first derivatives of Π_{HL} w.r.t. θ_{HL},θ_{L}, respectively. Then p^{L*} and p^{H*} can be obtained.

$$
\frac{\partial\Pi_{HL}}{\partial\theta_{HL}} = \theta_{L}\left(\alpha^{H}-\alpha^{L}\right)+\alpha^{L}-\alpha^{H}+\left(q^{H}-q^{L}\right)\left(1-2\theta_{HL}\right)
$$

$$
\begin{aligned}
\frac{\partial\Pi_{HL}}{\partial\theta_{L}} &= \left(q^{L}-\alpha^{H}\right)\left(1-\theta_{HL}\right)+\left(q^{L}-\alpha^{L}\right)\left(\theta_{HL}-2\theta_{L}\right)-\alpha^{L}+u_{0} \\
&= \theta_{HL}\left(\alpha^{H}-\alpha^{L}\right)-2\theta_{L}\left(q^{L}-\alpha^{L}\right)+q^{L}-\alpha^{H}-\alpha^{L}+u_{0}
\end{aligned}
$$

Alternatively, we can write the marginal consumer types as $\theta_{L}=\frac{p^{L}+u_{0}-\alpha^{L}}{q^{L}-\alpha^{L}}$, $\theta_{HL}=\frac{p^{H}-p^{L}-\left(\alpha^{H}-\alpha^{L}\right)\frac{q^{L}-p^{L}-u_{0}}{q^{L}-\alpha^{L}}}{q^{H}-q^{L}}$, and the total demand is $D=1-\theta_{L}$. With $0<\theta_{L}<1$ and $0<\theta_{H}<1$, we must have $p^{L}+u_{0}<q^{L}$ and $q^{L}>\alpha^{L}$. We can express the profit as

$$
\begin{aligned}
\Pi_{HL}\left(p^{L},p^{H}\right) &= p^{L}\left(\theta_{HL}-\theta_{L}\right)+p^{H}\left(1-\theta_{HL}\right) \\
&= p^{L}\left[\frac{\left(p^{H}-p^{L}\right)\left(q^{L}-\alpha^{L}\right)-\left(\alpha^{H}-\alpha^{L}\right)\left(q^{L}-p^{L}-u_{0}\right)}{\left(q^{H}-q^{L}\right)\left(q^{L}-\alpha^{L}\right)}-\frac{p^{L}+u_{0}-\alpha^{L}}{q^{L}-\alpha^{L}}\right] \\
&+ p^{H}\left[1-\frac{\left(p^{H}-p^{L}\right)\left(q^{L}-\alpha^{L}\right)-\left(\alpha^{H}-\alpha^{L}\right)\left(q^{L}-p^{L}-u_{0}\right)}{\left(q^{H}-q^{L}\right)\left(q^{L}-\alpha^{L}\right)}\right]
\end{aligned}
$$

When $0 \leq \frac{p^{L*}+u_0-\alpha^L}{q^L-\alpha^L} < \frac{p^{H*}-p^{L*}-\left(\alpha^H-\alpha^L\right)\frac{q^L-p^{L*}-u_0}{q^L-\alpha^L}}{q^H-q^L} < 1$, $0 < p^{L*} < p^{H*}$ and $\Pi^*_{HL} \geq \frac{\left(q^H-u_0\right)^2}{4(q^H-\alpha^H)}$, the firm would offer two products with positive prices and

$$\theta^*_L = \frac{\left(\alpha^H-\alpha^L\right)\left(\alpha^H-\alpha^L-q^H+q^L\right)+2\left(q^H-q^L\right)\left(\alpha^H+\alpha^L-u_0-q^L\right)}{\left(\alpha^H-\alpha^L\right)^2-4\left(q^H-q^L\right)\left(q^L-\alpha^L\right)}$$

$$\theta^*_{HL} = \frac{\left(\alpha^H-\alpha^L\right)^2+\alpha^L\left(u_0+2q^H-3q^L\right)+\alpha^H\left(q^L-u_0\right)-2q^L\left(q^H-q^L\right)}{\left(\alpha^H-\alpha^L\right)^2-4\left(q^H-q^L\right)\left(q^L-\alpha^L\right)}$$

$$\Pi^*_{HL} = \frac{\left(q^H-q^L\right)\left[\alpha^L\left(q^H-u_0\right)-\alpha^H\left(q^L-u_0\right)+\left(2u_0-q^H\right)q^L-u_0^2\right]}{\left(\alpha^H-\alpha^L\right)+4\left(\alpha^L-q^L\right)\left(q^H-q^L\right)}$$

When $\frac{\partial \Pi_{HL}}{\partial p^L}|_{p^{L*}=0,p^{H*}>0} \leq 0$, $0 < \frac{u_0-\alpha^L}{q^L-\alpha^L} < \frac{\left(q^H-q^L\right)q^L-\alpha^L\left(q^H-2q^L+u_0\right)+\alpha^H\left(u_0-q^L\right)}{2(q^H-q^L)(q^L-\alpha^L)} < 1$ and $\Pi^*_{HL} \geq \frac{\left(q^H-u_0\right)^2}{4(q^H-\alpha^H)}$, the firm offers two products with:

$p^{L*}=0$, $p^{H*} = \frac{q^L\left(\alpha^H-q^L\right)-\left(\alpha^H-\alpha^L\right)u_0}{2(q^L-\alpha^L)} + \frac{q^H}{2}$,

$\theta^*_L = \frac{u_0-\alpha^L}{q^L-\alpha^L}$,$\theta^*_{HL} = \frac{\left(q^H-q^L\right)q^L-\alpha^L\left(q^H-2q^L+u_0\right)+\alpha^H\left(u_0-q^L\right)}{2(q^H-q^L)(q^L-\alpha^L)}$,

$\Pi^*_{HL} = \frac{\left[\alpha^L\left(q^H-u_0\right)-q^L\left(q^H-q^L\right)+\alpha^H\left(u_0-q^L\right)\right]^2}{4(q^H-q^L)(q^L-\alpha^L)^2}$.

Otherwise the firm offers only the high-end product with $p^{H*} = \frac{q^H-u_0}{2}$,$\theta^*_H = \frac{q^H+u_0-2\alpha^H}{2(q^H-\alpha^H)}$, $\Pi^*_H = \frac{\left(q^H-u_0\right)^2}{4(q^H-\alpha^H)}$. □

Proof. ***Concavity of profit function in*** q^i**.**

When q^i is an endogenous decision, we prove that, if $V^i(\theta, D)$ is concave in q^i and $C\left(q^i\right)$ is convex in q^i for $i \in \{H, L\}$, the profit function is concave in q^i .

First, we look at the case where only the high-end product is provided. By Lemma 1, we have θ_H defined by $V^H(\theta_H, D) - p^H = u_0$, so $p^H = V^H(\theta_H, D) - u_0$.

$$\Pi_H = \left[p^H - C\left(q^H\right)\right] \int\limits_{\theta_H}^{1} f(\theta)\, d\theta$$
$$= \left[V^H(\theta_H, D) - u_0 - C\left(q^i\right)\right] \int\limits_{\theta_H}^{1} f(\theta)\, d\theta$$

$$\frac{\partial \Pi_H}{\partial q^H} = \left[\frac{\partial V^H(\theta_H, D)}{\partial q^H} - \frac{\partial C\left(q^i\right)}{\partial q^H}\right] \int\limits_{\theta_H}^{1} f(\theta)\, d\theta$$

$$\frac{\partial^2 \Pi_H}{\partial \left(q^H\right)^2} = \left[\frac{\partial^2 V^H(\theta_H, D)}{\partial \left(q^H\right)^2} - \frac{\partial^2 C\left(q^i\right)}{\partial \left(q^H\right)^2}\right] \int\limits_{\theta_H}^{1} f(\theta)\, d\theta$$

Therefore, when $V^i(\theta, D)$ is concave in q^i, and $C\left(q^i\right)$ is convex in q^i for $i \in \{H, L\}$, we have $\frac{\partial^2 \Pi_H}{\partial (q^H)^2} < 0$; thus, Π_H is concave in q^H.

Following the same logic, we can get Π_{HL} is concave in q^H and q^L. □

Proof. ***Proposition 3.***

In conventional product line design

$$\Pi_{HL} = (1 - \theta_{HL})\left[p^H - c\left(q^H\right)^2\right] + (\theta_{HL} - \theta_L)\left[p^L - c\left(q^L\right)^2\right]$$
$$s.t. p^L = \theta_L q^L + \alpha^L (1 - \theta_L) - u_0$$
$$p^H = \theta_{HL}\left(q^H - q^L\right) + \left(\alpha^H - \alpha^L\right)(1 - \theta_L) + p^L$$
$$0 < \theta_L < \theta_{HL} < 1$$

Because Π_{HL} is concave in q^i, with first order conditions w.r.t. q^i, we have:

$$\begin{cases} q^{H*} = \frac{\theta_{HL}}{2c} \\ q^{L*} = \frac{\theta_{HL}+\theta_L-1}{2c} \end{cases}$$

Substituting q^{H*} and q^{L*} into Π_{HL}, we can derive the optimal desicision of $\theta^*_{HL}, \theta^*_L$:

$$\begin{cases} \theta^*_{HL} = \frac{1}{15}\left(11 + 8\alpha^L c - \sqrt{1 - 64\alpha^L c + 64\left(\alpha^L\right)^2 c^2 + 60cu_0}\right) \\ \theta^*_L = \frac{1}{15}\left(7 + 16\alpha^L c - 2\sqrt{1 - 64\alpha^L c + 64\left(\alpha^L\right)^2 c^2 + 60cu_0}\right) \end{cases}$$

or

$$\begin{cases} \theta^*_{HL} = \frac{1}{15}\left(11 + 8\alpha^L c + \sqrt{1 - 64\alpha^L c + 64\left(\alpha^L\right)^2 c^2 + 60cu_0}\right) \\ \theta^*_L = \frac{1}{15}\left(7 + 16\alpha^L c + 2\sqrt{1 - 64\alpha^L c + 64\left(\alpha^L\right)^2 c^2 + 60cu_0}\right) \end{cases}$$

subject to $0 < \theta^*_L < \theta^*_{HL} < 1$. Substituting into $q^{H*}$and q^{L*}, we can get $\frac{\partial q^{L*}}{\partial \alpha^L} > 0$.

When freemium is optimal, we have

$$\begin{aligned} &\Pi_F = (1-\theta_{HL})\left[p^H - c\left(q^H\right)^2\right] + (\theta_{HL} - \theta_L)\left[-c\left(q^L\right)^2\right] \\ &s.t. p^L = 0 = \theta_L q^L + \alpha^L(1-\theta_L) - u_0 \\ &\quad p^H = \theta_{HL}\left(q^H - q^L\right) + \left(\alpha^H - \alpha^L\right)(1-\theta_L) \\ &\quad 0 < \theta_L < \theta_{HL} < 1 \end{aligned}$$

As Π_F is concave in q^i. With first order conditions w.r.t q^H, we have $q^{H*} = \frac{\theta_{HL}}{2c}$. Taking derivative w.r.t q^L, we have $\frac{\partial \Pi_F}{\partial q^L} < 0$ always holds.

$$\begin{cases} q^{H*} = \frac{\theta_{HL}}{2c} \\ q^{L*} = \frac{u_0 - \alpha^L(1-\theta_L)}{\theta_L} \end{cases} \tag{8}$$

With (8) , we have

$$\frac{\partial q^{L*}}{\partial \alpha^L} = \frac{\theta_L - 1}{\theta_L} < 0$$

$\square$

Proof. ***Proposition 4.***

With $\alpha^H = \alpha^L = \alpha$, we now have

$$V^L(\theta, \alpha, D) = \theta \left(q^L + \alpha D\right),$$
$$V^H(\theta, \alpha, D) = \theta \left(q^H + \alpha D\right),$$

where D is the total demand of all offered products. As is typical in games with network effects, multiple equilibria may exist in the second stage. We seek the Nash Equilibrium that is Pareto dominant. More specifically, when network effects are intermediate, there exist multiple equilibria where all, some, or none of the consumers adopt the products. When consumers do not adopt, the products do not generate sufficient network effects and thus non-adoption becomes self-fulfilling. This coordination failure is classic in models with network effects. Clearly, the equilibrium wherein all users adopt generates (weakly) higher surplus for all parties. Thus, we select that equilibrium whenever it exists. For the proof of Proposition 4, two cases are analyzed below.

Case 1: Sell to both segments with $(q^H, \alpha, p^H), (q^L, \alpha, p^L)$

When network effects are present, the binding constraints in the firm's optimization problem continue to be the low-end consumers' IR constraint and the high-end consumers' IC constraint. The optimal prices satisfy:

$$p^L = \theta_L(q^L + \alpha) - u_0$$

$$p^H = \theta_H(q^H - q^L) + \theta_L(q^L + \alpha) - u_0$$

We bound q^L above zero in the following analysis. The optimal qualities can therefore be determined by a profit maximizing problem wherein:

$$\Pi_{HL} = \lambda \left[p^H - c\left(q^H\right)^2 - s\alpha^{\mathbf{2}}\right] + (1-\lambda)\left[p^L - c\left(q^L\right)^2 - s\alpha^{\mathbf{2}}\right] \ s.t.\ p^L{\geq}0,\ q^L{\geq}0$$

Using Lagrangian method, the optimal menu is

$$q^{H*} = \frac{\theta_H}{2c}$$

$$q^{L*} = \begin{cases} \frac{\theta_L - \theta_H\lambda}{2c(1-\lambda)} & ,\lambda\theta_H\theta_L{\leq}\theta_L^2\ \&\ \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{\theta_L^2(c-c\lambda+s)} - 1 \leq 0 \\ 0 & ,\lambda\theta_H\theta_L > 2su_0 > \theta_L^2 \\ 0 & ,\lambda\theta_H\theta_L > \theta_L^2 \geq 2su_0 \\ \frac{2su_0\theta_L - \theta_H\lambda}{2\theta_L(c-c\lambda+s)} & ,2su_0 > \lambda\theta_H\theta_L \& 2su_0 > \theta_L^2 \& \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{c-c\lambda+\theta_L^2 s} - 1 > 0 \end{cases}$$

$$\alpha^* = \begin{cases} \frac{\theta_L}{2s} & ,\lambda\theta_H\theta_L{\leq}\theta_L^2\ \&\ \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{\theta_L^2(c-c\lambda+s)} - 1 \leq 0 \\ u_0 & ,\lambda\theta_H\theta_L > 2su_0 > \theta_L^2 \\ \frac{\theta_L}{2s} & ,\lambda\theta_H\theta_L > \theta_L^2 \geq 2su_0 \\ \frac{\theta_H\theta_L\lambda+2cu_0(1-\lambda)}{2\theta_L(c-c\lambda+s)} & ,2su_0 > \lambda\theta_H\theta_L \& 2su_0 > \theta_L^2 \& \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{c-c\lambda+\theta_L^2 s} - 1 > 0 \end{cases}$$

$$p^{L*} = \begin{cases} \frac{\theta_L(\theta_L-\theta_H\lambda)}{2c(1-\lambda)} + \frac{\theta_L^2}{2s} - u_0 & ,\lambda\theta_H\theta_L{\leq}\theta_L^2\ \&\ \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{\theta_L^2(c-c\lambda+s)} - 1 \leq 0 \\ 0 & ,\lambda\theta_H\theta_L > 2su_0 > \theta_L^2 \\ \frac{\theta_L^2}{2s} - u_0 & ,\lambda\theta_H\theta_L > \theta_L^2 \geq 2su_0 \\ 0 & ,2su_0 > \lambda\theta_H\theta_L \& 2su_0 > \theta_L^2 \& \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{c-c\lambda+\theta_L^2 s} - 1 > 0 \end{cases}$$

$$p^{H*} = \begin{cases} \frac{\theta_H^2+\theta_L^2-\theta_H\theta_L(1+\lambda)}{2c(1-\lambda)} + \frac{1}{2s} - u_0 & ,\lambda\theta_H\theta_L{\leq}\theta_L^2\ \&\ \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{\theta_L^2(c-c\lambda+s)} - 1 \leq 0 \\ \frac{\theta_H^2}{2c} & ,\lambda\theta_H\theta_L > 2su_0 > \theta_L^2 \\ \frac{\theta_H^2}{2c} + \frac{\theta_L^2}{2s} - u_0 & ,\lambda\theta_H\theta_L > \theta_L^2 \geq 2su_0 \\ \theta_H\left(\frac{\theta_H}{2c} - \frac{2su_0-\theta_H\theta_L\lambda}{2\theta_L(c-c\lambda+s)}\right) & ,2su_0 > \lambda\theta_H\theta_L \& 2su_0 > \theta_L^2 \& \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{c-c\lambda+\theta_L^2 s} - 1 > 0 \end{cases}$$

$$\Pi_{HL}^* = \begin{cases} \frac{\theta_L^2+\theta_H^2\lambda-2\theta_H\theta_L\lambda}{4c(1-\lambda)} + \frac{\theta_L^2}{4s} - u_0 & ,\lambda\theta_H\theta_L \leq \theta_L^2 \ \& \ \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{\theta_L^2(c-c\lambda+s)} - 1 \leq 0 \\ \frac{\theta_H^2\lambda}{4c} - \frac{su_0^2}{\theta_L^2} & ,\lambda\theta_H\theta_L > 2su_0 > \theta_L^2 \\ \frac{\theta_H^2\lambda}{4c} + \frac{\theta_L^2}{4s} - u_0 & ,\lambda\theta_H\theta_L > \theta_L^2 \geq 2su_0 \\ \frac{\theta_H^2\theta_L^2\lambda(c+s)-4\theta_H\theta_L scu_0\lambda-4sc^2u_0^2(1-\lambda)}{4\theta_L^2 c(c+s-c\lambda)} & ,2su_0 > \lambda\theta_H\theta_L \& 2su_0 > \theta_L^2 \& \frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{c-c\lambda+\theta_L^2 s} - 1 > 0 \end{cases}$$

Case 2: Sell to θ_H-segment only with(q, α, p)

Binding condition $p = \theta_H(q+\alpha\lambda) - u_0$. Firm's profit is given by $\Pi_H = \lambda(p - cq^2 - s\alpha^2)$. Solving the optimization problem, we obtain:

$$p^* = \frac{\theta_H^2}{2c} + \frac{\lambda^2\theta_H^2}{2s} - u_0, \ q^* = \frac{\theta_H}{2c}, \ \alpha^* = \frac{\lambda\theta_H}{2s}$$

$$\Pi_H^* = \lambda\left(\frac{\theta_H^2}{4c} + \frac{\lambda^2\theta_H^2}{4s} - u_0\right)$$

Freemium is optimal if and only if (IFF) the conditions $\Pi_{HL}^* \geq \Pi_H^*, \ p^{L*} = 0$ are satisfied.

Below, we prove the above conditions cannot hold simultaneously. According to the results obtained by using the Lagrangian method, we discuss by parameter ranges where p^{L*} may possibly be zero.

(1)When $\lambda\theta_H\theta_L > 2su_0 > \theta_L^2$.

We have

$$\begin{aligned} \Pi_H^* - \Pi_{HL}^* &= \lambda\left(\frac{\theta_H^2\lambda^2}{4s} - u_0\right) + \frac{su_0^2}{\theta_L^2} \\ &> \lambda\left(u_0\frac{su_0}{\theta_L^2} - u_0\right) + \frac{u_0}{2} \\ &> u_0\left(\frac{1}{2} - \frac{\lambda}{2}\right) > 0 \end{aligned}$$

Therefore, $\Pi_{HL}^* < \Pi_H^*$ always holds when $\lambda\theta_H\theta_L > 2su_0 > \theta_L^2$; thus, the firm prefers to offer only the high-end product, and freemium cannot emerge.

(2) When $2su_0 > \lambda\theta_H\theta_L$, $2su_0 > \theta_L^2$ and $\frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{c-c\lambda+\theta_L^2 s} - 1 > 0$.

Following the same logic as in (1), we prove $\Pi_H^* > \Pi_{HL}^*$ always holds as long as the high-type consumer finds the low-end product worth trying at price zero, i.e., $\theta_L(q^L+\lambda\alpha)-u_0 \geq 0$.

$$\begin{aligned}\Pi^*_{HL} &= \frac{\theta_H^2\theta_L^2\lambda(c+s) - 4\theta_H\theta_L scu_0\lambda - 4sc^2u_0^2(1-\lambda)}{4\theta_L^2 c(c+s-c\lambda)} \\ &= \frac{\lambda\theta_H^2}{4c} + \frac{\theta_H^2\lambda^2}{4(c+s-c\lambda)} - \frac{su_0[\theta_H\theta_L\lambda + cu_0(1-\lambda)]}{\theta_L^2(c+s-c\lambda)}\end{aligned}$$

$$\Pi^*_H = \frac{\lambda\theta_H^2}{4c} + \lambda\left(\frac{\lambda^2\theta_H^2}{4s} - u_0\right)$$

Let $F_{HL} = \frac{\theta_H^2\lambda^2}{4(c+s-c\lambda)} - \frac{su_0[\theta_H\theta_L\lambda+cu_0(1-\lambda)]}{\theta_L^2(c+s-c\lambda)}$, $F_H = \lambda\left(\frac{\lambda^2\theta_H^2}{4s} - u_0\right)$. The IFF condition for $\Pi^*_H > \Pi^*_{HL}$ is $F_H > F_{HL}$. We proceed by proving that, under the conditon $p^{L*} < 0$, $F_H > F_{HL(max)}$ holds; thus, $F_H > F_{HL}$ and $\Pi^*_H > \Pi^*_{HL}$.

Below we prove F_{HL} is decreasing in c; thus, $\underset{c>0}{sup} F_{HL} = \underset{c\to 0}{lim} F_{HL} = F_{HL}|_{c=0} = \frac{\theta_H^2\lambda^2}{4s} - \frac{u_0\theta_H\lambda}{\theta_L}$. Therefore, a sufficient condition for $F_H > F_{HL}$ is $F_H > \underset{c>0}{sup} F_{HL}$, or equivalently, $F_H > F_{HL}|_{c=0}$. (Notice F_H is not a function of c.)

$$\begin{aligned}\frac{\partial F_{HL}}{\partial c} &= -\frac{\theta_H^2\theta_L^2\lambda(1-\lambda)}{4\theta_L^2(c+s-c\lambda)} - \frac{4su_0(su_0-\theta_H\theta_L\lambda)(1-\lambda)}{4\theta_L^2(c+s-c\lambda)} \\ &= -\frac{(1-\lambda)[4su_0(su_0-\theta_H\theta_L\lambda)+\theta_H^2\theta_L^2\lambda]}{4\theta_L^2(c-c\lambda+s)}\end{aligned}$$

We have $4su_0(su_0-\theta_H\theta_L\lambda)+\theta_H^2\theta_L^2\lambda > 4\frac{\lambda\theta_H\theta_L}{2}(\frac{\lambda\theta_H\theta_L}{2} - \theta_H\theta_L\lambda) + \theta_H^2\theta_L^2\lambda = -\theta_H^2\theta_L^2\lambda^2 + \theta_H^2\theta_L^2\lambda > 0$, hence $\frac{\partial F_{HL}}{\partial c} < 0$, and F_{HL} is decreasing in c.

Next we prove $F_H > F_{HL}|_{c=0}$. When c approaches 0, the condition $\frac{s[\theta_H\theta_L\lambda+2cu_0(1-\lambda)]}{\theta_L^2(c-c\lambda+s)} - 1 > 0$ implies $\lambda > \frac{\theta_L}{\theta_H}$. We first prove $F_{HL}|_{c=0} < 0$ for all $\lambda > \frac{\theta_L}{\theta_H}$.

At $\lambda > \frac{\theta_L}{\theta_H}$, we have

$$F_{HL}|_{c=0,\lambda=\frac{\theta_L}{\theta_H}} = \frac{\theta_L^2}{4s} - u_0 < \frac{2su_0}{4s} - u_0 = -\frac{u_0}{2} \leq 0$$

and

$$\frac{\partial F_{HL}|_{c=0}}{\partial \lambda} = \frac{\theta_H^2 \lambda}{2s} - \frac{u_0 \theta_H}{\theta_L} < \frac{2su_0/\theta_L}{2s}\theta_H - \frac{u_0\theta_H}{\theta_L} = 0$$

Therefore, when $\lambda > \frac{\theta_L}{\theta_H}$, we always have $F_{HL}|_{c=0} < 0$.

We also have

$$\begin{aligned} F_H &= \lambda \left(\frac{\lambda^2 \theta_H^2}{4s} - u_0 \right) \\ &> \lambda \left(\frac{\lambda^2 \theta_H^2}{4s} - u_0 \frac{\theta_H}{\theta_L} \lambda \right) \\ &= \lambda F_{HL}|_{c=0} \\ &> F_{HL}|_{c=0} \text{ (since } F_{HL}|_{c=0} < 0) \end{aligned}$$

So we have proved $F_H > F_{HL}|_{c=0}$; hence, $F_H > F_{HL}$ is also proved. □

Proof. ***Proposition 5.***

The proof follows similar logic as in Proposition 4. With asymmetric network effects, consumer valuations of the products are:

$$V^L(\theta, \alpha^L, D) = \theta(q^L + \alpha^L D),$$
$$V^H(\theta, \alpha^H, D) = \theta(q^H + \alpha^H D).$$

Case 1: Sell to both segments with $(q^H, \alpha^H, p^H), (q^L, \alpha^L, p^L)$.

The optimal prices satisfy:

$$p^L = \theta_L(q^L + \alpha^L) - u_0$$
$$p^H = \theta_H(q^H - q^L + \alpha^H - \alpha^L) + \theta_L(q^L + \alpha^L) - u_0$$

The optimal qualities can therefore be determined by a profit maximizing problem wherein:

$$\Pi_{HL} == \lambda\left[p^H - c\left(q^H\right)^2 - s\left(\alpha^H\right)^2\right] + (1-\lambda)\left[p^L - c\left(q^L\right)^2 - s\left(\alpha^L\right)^2\right] \ s.t.\ p^L{\geq}0,\ q^L{\geq}0$$

The optimal quality levels remain the same as in the no-network-effects scenario, namely

$q^{H*} = \frac{\theta_H}{2c}$

$q^{L*} = \begin{cases} \frac{\theta_L - \theta_H\lambda}{2c(1-\lambda)} & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 \leq 0 \\ \frac{su_0}{\theta_L(c+s)} & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 > 0 \end{cases}$

$\alpha^{H*} = \frac{\theta_H}{2s}$

$\alpha^{L*} = \begin{cases} \frac{\theta_L - \theta_H\lambda}{2s(1-\lambda)} & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 \leq 0 \\ \frac{cu_0}{\theta_L(c+s)} & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 > 0 \end{cases}$

$p^{L*} = \begin{cases} \frac{\theta_L(\theta_L - \theta_H\lambda)}{2(1-\lambda)}\left(\frac{1}{c} + \frac{1}{s}\right) - u_0 & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 \leq 0 \\ 0 & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 > 0 \end{cases}$

$p^{H*} = \begin{cases} \frac{\theta_H^2 + \theta_L^2 - (1+\lambda)\theta_L\theta_H}{2(1-\lambda)}\left(\frac{1}{c} + \frac{1}{s}\right) - u_0 & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 \leq 0 \\ \frac{\theta_H^2}{2}\left(\frac{1}{c} + \frac{1}{s}\right) - \frac{u_0\theta_H}{\theta_L} & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 > 0 \end{cases}$

The optimal profit is:

$\Pi_{HL}^* = \begin{cases} \frac{\theta_H^2 + \theta_L^2\lambda - 2\lambda\theta_L\theta_H}{4(1-\lambda)}\left(\frac{1}{c} + \frac{1}{s}\right) - u_0 & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 \leq 0 \\ \frac{\theta_H^2\lambda(c+s)}{4cs} - \frac{\theta_H\lambda u_0}{\theta_L} - \frac{csu_0^2(1-\lambda)}{\theta_L^2(c+s)} & , \frac{2csu_0(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 > 0 \end{cases}$

Case 2: Sell to θ_H-segment only

Here the situation is the same as in case 2 in the symmetric network effects senario, where $p = \theta_H(q + \alpha\lambda) - u_0$ and $\Pi_H = \lambda(p - cq^2 - s\alpha^2)$:

$$q^* = \frac{\theta_H}{2c},\ \alpha^* = \frac{\lambda\theta_H}{2s},$$

$$\Pi_H^* = \lambda\left(\frac{\theta_H^2}{4c} + \frac{\lambda^2\theta_H^2}{4s} - u_0\right)$$

For freemium to be optimal, the conditions $\Pi_{HL}^* \geq \Pi_H^*,\ p^{L*} = 0$ have to be satisfied.

Under asymmetric network effects, freemium equilibrium exists, when (s, c, u_0, λ) satisfy the conditions below:

$$\frac{\theta_H^2}{4s}(\lambda - \lambda^3) + \lambda u_0 \left(1 - \frac{\theta_H}{\theta_L}\right) - \frac{cs(1-\lambda)u_0^2}{\theta_L^2(c+s)} \geq 0$$
$$\frac{\theta_H{}^{\mathbf{2}}\lambda(c+s)}{4sc} - \frac{cs(1-\lambda)u_0^2}{\theta_L^2(c+s)} - \frac{\theta_H \lambda u_0}{\theta_L} > 0$$
$$\frac{2csu_0\,(1-\lambda)}{\theta_L^2(c+s)} + \frac{\lambda\theta_H}{\theta_L} - 1 > 0$$

It can be seen that the above inequalities define a non-empty set. Corollary 4 follows straightforwardly according to the above results. □

Appendix B

In the analysis in section 1.5.2, we alluded to the possibility that the assumption "the high-type consumers have positive valuation for the low-end product at price zero" may not hold. This means that the highest-type consumers will not find the low-end product worth trying even if it is offered at zero price, which is very unlikely for successful freemium products. In this section, we provide a detailed analysis for this exceptional case and show that even if the assumption is violated, our results still hold true unless the cost function has some special form. With some special cost function and the assumption violated (i.e., $V^L(\theta_H, \alpha, \lambda) < 0$), raising the low-end product's price above zero will violate the high type's IR instead of IC constraint, then it is possible that freemium would emerge under uniform network effects.

As in the baseline model, the firm can either serve the high type consumers with menu (p, q, α_1), getting profit Π; or serve both high and low segments, with menu $(p^H, q^H, \alpha_2), (p^L, q^L, \alpha_2)$, getting profit Π_{HL}. Freemium is a special case of the second strategy, where $p^L = 0$ and profit Π_F. We focus on necessary conditions for freemium to be optimal.

Let $C_q(q)$, $C_\alpha(\alpha)$ be the marginal cost for offering one product of quality q and network benefit α.

(1)When $\lambda > \frac{\theta_L}{\theta_H}$.

By checking only p and p^H, we can see that the firm prefers offering only the high-end product to adopting freemium strategy. More specifically, consider p^{H*}:

$$
\begin{aligned}
p^{H*} &= \theta_H(q^{H*} - q^{L*}) + \theta_L(q_L^* + \alpha_2^*) - u_0 \\
&< \theta_H(q^{H*} - q^{L*}) + \theta_H q_L^* + \theta_L \alpha_2^* - u_0 \\
&= \theta_H q^{H*} + \theta_L \alpha_2^* - u_0
\end{aligned}
$$

Now let's consider p, where we let the firm sets $\alpha_1 = \alpha_2^*$:

$$
\begin{aligned}
p &= \theta_H(q^* + \lambda\alpha_2^*) - u_0 \ (IR\ constraint) \\
&= \theta_H q^{H*} + \theta_H \lambda \alpha_2^* - u_0 \ (since\ q^* = q^{H*}) \\
&> \theta_H q^{H*} + \theta_L \alpha_2^* - u_0 \ (since\ \lambda > \frac{\theta_L}{\theta_H}) \\
&> p^{H*}
\end{aligned}
$$

Therefore, compared to serving only the high-type consumers, it is never optimal to pursue the freemium strategy, because the firm will get less profit on the high end while subsidizing the low-end segment. In fact, when selling only to the high-type customers, the firm can even increase profit by adopting an optimal α_1^*, making $\Pi^* > \Pi_{HL}^*$ always hold under $\lambda > \frac{\theta_L}{\theta_H}$.

(2)When $\lambda \leq \frac{\theta_L}{\theta_H}$.

First, consider serving only the high end. We have:

$$
\begin{aligned}
p &= \theta_H(q + \lambda\alpha_1) - u_0 \\
\Pi &= \lambda[p - C_q(q) - C_\alpha(\alpha_1)] \\
&= \lambda[\theta_H(q + \lambda\alpha_1) - u_0 - C_q(q) - C_\alpha(\alpha_1)]
\end{aligned}
$$

To maximize Π, we have:

$$
\begin{aligned}
C_q{'}(q^*) &= \theta_H \\
C_\alpha{'}(\alpha_1^*) &= \lambda\theta_H
\end{aligned}
$$

for $\Pi^* > 0$, we have$u_0 < \theta_H(q^* + \lambda\alpha_1^*) - C_q(q^*) - C_\alpha(\alpha_1^*)$.

Then, when the firm serves both segments, we have:

$$
\begin{aligned}
p^H &= \theta_H(q^H - q^L) + \theta_L(q^L + \alpha_2) - u_0 \\
p^L &= \theta_L(q^L + \alpha_2) - u_0 \\
\Pi_{HL} &= \lambda[p^H - C_q(q^H)] + (1-\lambda)[p^L - C_q(q^L)] - C_\alpha(\alpha_2)
\end{aligned}
$$

To maximize Π_{HL}, we have:

$$
\begin{aligned}
C_q{'}(q^{H*}) &= \theta_H \\
C_q{'}(q^{L*}) &= \frac{\theta_L - \lambda\theta_H}{1-\lambda} \\
C_\alpha{'}(\alpha_2^*) &= \theta_L
\end{aligned}
$$

Notice the necessary condition for freemium is $p^{L*} \leq 0$, where $p^{L*} = \theta_L(q^{L*} + \alpha_2^*) - u_0$. Under this condition, we check whether the firm would like to induce more cost (by offering higher-than-optimal $q^L + \alpha_2$ to make $p^L = 0$) to have the low-type consumers on board.

Supposing the firm adopts freemium, we need to compare Π^* and Π_F, where the firm set q^L and α_2 such that $p_L = \theta_L(q^L + \alpha_2) - u_0 = 0$ and $q^{H*} > q^L \geq q^{L*}$. Notice that $q^* = q^{H*}$, and q^{H*} is always equal to the efficient quality.

Assuming a convex cost function, with $u_0 < \theta_H(q^* + \lambda\alpha_1^*) - C_q(q^*) - C_\alpha(\alpha_1^*)$, we have:

$$
\begin{aligned}
\Delta\Pi &= \Pi_F^* - \Pi^* \\
&= \lambda(u_0 - \alpha_1^*\lambda\theta_H - \theta_H q^L) - (1-\lambda)C_q(q^L) + \lambda C_\alpha(\alpha_1^*) - C_\alpha(\alpha_2) \\
&< \lambda(\theta_H q^{H*} - C_q(q^*) - \theta_H q^L) - (1-\lambda)C_q(q^L) - C_\alpha(\alpha_2) \\
&< \lambda[C_q{'}(q^{H*})(q^{H*} - q^L) - C_q(q^{H*})] - (1-\lambda)C_q(q^L)
\end{aligned}
$$

We can see that $\Delta\Pi < 0$ unless $C_q(\cdot)$ is very steep and skewed towards zero, specifically, $C_q\prime(q^{H*}) > \left(\frac{1-\lambda}{\lambda}\right)\frac{C_q(q^L)}{q^{H*}-q^L} + \frac{C_q(q^{H*})}{q^{H*}-q^L}$.

Above we analyzed the case where network effect is endogenous. As can be easily seen from the analysis, exactly the same conclusion can be reached for the case where network effects are exogenously given, no matter how large α is.

Chapter 2
Hype News May Drive Real News: The Oz Effect in Healthcare

2.1 Introduction

As Reuters reported,[1] health information is one of the most frequently sought information on the internet. On an average, 53% of American people search for health information online.[2] While publicly available health information can sometimes help us better manage our health, alleviate concerns about our wellness, or even avoid some hospital visits, it can also be the source for incorrect or misleading information. In 2016, the sensational tragedy of Zexi Wei in China drew unprecedented public attention to the issue of credibility of online healthcare information.[3] Baidu.com was scolded as an accessory to murder because it did not pull misleading medical information that recommended unproven methods to treat a rare form of cancer from its search results. Given the enormous growth of medical information online, especially from less credible sources, there may be a serious risk of erroneous information or exaggerated information influencing medical treatments. This problem is especially acute in the case of serious public health challenges such as obesity that seem to defy most treatments.

In spite of the importance of public health information, how information from different sources affects people's healthcare choices is poorly understood. This paper aims to fill this important research gap, by studying the effect of public information on the demand for over-the-counter (OTC) weight loss products. For OTC healthcare products, consumers' purchasing decisions are especially prone to be affected by the publicly available information because consumers are free to choose what information to believe and do not need a physician's prescription. A good representative of OTC healthcare products is weight loss product. In the US, more that 70% of adults aged 20 and over are overweight, including obesity,[4] and more than 30% of people who made weight loss attempts used OTC weight loss products (Eisenberg, 2008). And globally, the revenue for weight loss and weight management market is expected to increase from $15.9 billion in 2016 to $22.9 billion by 2025.[5] Given the importance of the obesity issue and the

---

[1] Reuters. Consumer-targeted internet investment: online strategies to improve patient care and product positioning. Reuters

[2] Pew Internet American Life Project. Internet visits soaring. Health Management Technology 2003; 24: 2–8.

[3] Wikipedia (2017). *Death of Wei Zexi*. Retrieved from https://en.wikipedia.org/wiki/Death_of_Wei_Zexi

[4] Center for Disease Control and Prevention (2013-2014). Retrieved from https://www.cdc.gov/nchs/fastats/obesity-overweight.htm

[5] PRNewswire (2017). Retrieved from https://www.prnewswire.com/news-releases/global-weight-loss-and-weight-management-diet-market-is-expected-to-grow-to-229188-by-2025-300488240.html

enormous market value (Khan et al., 2016), we choose to focus on weight loss products in this study. We examine almost all public health information sources, including online customer reviews, scholarly journals, newspapers, magazines, dissertations, working papers, market reports, and medical TV shows. Although other information sources, such as doctors' advice, and word of mouth from friends and families may also influence consumer healthcare choices, their effects are idiosyncratic, whereas public information influences the mass market collectively. Considering the inevitable role public information plays in shaping the aggregate demand for weight loss products, we focus on public information sources in this study.

To have a comprehensive understanding the effect of public information, we decompose each piece of information to two dimensions: (i) the credibility of the information source and (ii) the content of the information.

First of all, prior literature suggests that source credibility plays an indispensable role in affecting consumer's healthcare choices (e.g., Hu & Sundar, 2010; Bates et al., 2006). According to credibility levels, we categorize the different sources of public health information into the following four groups:

1. Peer-reviewed scientific research articles. Compared to other sources, research articles are highly credible, because of the serious review process and expertise of the medical researchers.

2. User-generated product reviews. On the one hand, customer reviews can convey first hand evaluation of the healthcare product/service; thus, they can be very credible and reliable. On the other hand, reviews can become problematic because a customer may not know how to give an unbiased evaluation. Moreover, reviews can be deceptive if they are manipulated by the product/service provider.

3. News articles (including those from newspapers, magazines, and news radio transcripts) written by professional journalists or reporters. News articles may not be highly credible because news reporters or journalists do not always write with evidence that can stand the test of time, as they may lack the expertise and time to mine the truth from all information they collect.

4. TV shows launched or participated in by doctors. A good representative of this source is *The Dr. Oz Show*.[6] Dr. Oz is a surgeon and professor at the department of surgery of Columbia University, and a television personality. He launched a daily show called *The Dr. Oz Show*, which focuses on medical issues and personal health. Products mentioned by Dr. Oz on the show often see sales skyrocketing afterward. This phenomenon is termed the "Oz effect."[7] Despite

[6] For an example of The Dr. Oz Show, please see *http://www.doctoroz.com/episode/newest-health-food-sensation-revealed*

[7] Walton, A. (2011). *The Oz Effect: Medicine or Marketing?* Retrieved from https://www.forbes.com/sites/alicegwalton/2011/06/06/the-oz-effect-medicine-or-marketing/#3e88143d3233

the success of his show,[8] he has been criticized by physicians and government officials for giving non-scientific advice. In 2014, The Federal Trade Commission filed a lawsuit against several products he peddled on his show, and he was scolded during the congressional hearing. There are also other TV shows (e.g., The Doctors[9]) that mentioned the products in our data. Given *The Dr. Oz Show* supplies the majority of this type of information according to the data, we categorize all these shows and call them *The Dr. Oz Show* for ease of reference.

Second, we differentiate information sources by content language features. For example, a research article is formal, neutral, and sometimes hard to understand, and it rarely mentions price information. In contrast, an online customer review is typically casual, emotional, and easy to understand. We seek to understand precisely what kinds of content impact consumer choices. To extract textual content features for each information, we employ both traditional machine learning and state-of-the-art deep learning natural language processing (henceforth "NLP") techniques. Among the textual features, we include not only informative language features such as product price and performance, but also persuasive aspects such as humor and subjectivity. We analyze thirteen textual features to understand whether and to what extent these features play a significant role in consumers' healthcare choices.

Our findings imply that *The Dr. Oz Show*, which is believed to be highly influential and viewed as hyped information in consumers' choice of healthcare products, does not significantly affect the sales of the recommended products. Instead, the sales boost of some controversial products may be caused not by Dr. Oz, but by public information from other sources. Specifically, news articles, as a fairly common media channel, are found to significantly affect the general public's choice in the healthcare domain. Moreover, the availability and language features of the most credible information—scientific research articles—play an important role in steering consumers' choices. Therefore, we propose an alternative explanation of the "Oz effect." That is, instead of affecting sales directly, the show has an indirect impact on consumers' purchasing decisions through stimulating more related "real" news articles.

The Dr. Oz Show serves as an example to demonstrate a general implication that, healthcare information from celebrity doctors may initiate a media-hype (Vasterman 2005; Zuckerman 2003), creating news waves on the recommended healthcare products. From a public policy perspective, while exaggerated information may cause attention, consumers rely on searching through credible sources to make decisions. This rational behavior offers a measure of comfort that mere puffery does not excessively influence decision on healthcare. An unexpected market reaction also contributes to this outcome. When hype news happens, it seems to drive real news. While Google, Baidu and Facebook are fighting fake news; our finding suggests that

[8] Dr. Oz won Daytime Emmy Award in 2010, 2011, 2012, 2013, and 2016.

[9] Please see https://www.thedoctorstv.com/ for its official website.

credible information sources can quickly correct fake news. Credible sources respond to the hype news by generating useful content. This mitigates and even negates the effect of hype news.

Above and beyond the information intensity, we find that content features also play a significant role in explaining consumer demand. Several interesting results emerge from our analysis. We find that the effect of credible sources can be further enhanced by lowering the complexity, ambiguity or emotion contained in research articles. For product reviews by consumers, we find that rating has a positive effect but sentiment has a negative effect on sales. Furthermore, we discover that a product's sales increase when the language used in the customer reviews is more positive, ambiguous and subjective. Our analysis also implies that mentioning a product's name or performance in news articles or product reviews has a significant positive effect on its sales.

In summary, this paper makes a substantive contribution to provide concrete empirical evidence of how information from various public sources affects consumer healthcare product demand. Methodologically, this is one of the first few papers that use deep learning methods to quantify content information from unstructured data to aid marketing research. Managerially, our findings about the content features provide guidance on how to design language so that the advertising efforts through multi-channel media can be most effective. The results also have important public policy implications on how to potentially regulate media content in the healthcare domain, in order to protect consumers from harmful health information.

We organize the rest of the paper as follows. In Section 2, the related literature and the contribution of the present paper are discussed. Section 3 describes the data and model-free evidence of information's impact on consumer choices. The methodology, including both NLP models and econometric analysis, are presented in Section 4. Section 5 discusses the results. Finally, in Section 6, we conclude with managerial and public policy implications, as well as limitations.

2.2 Literature review

To our knowledge, the work most related to ours is Ching et al. (2016), which investigates the impact of publicity on demand for Anti-Cholesterol drugs. As these drugs require a prescription, the impact of news articles must be filtered through physicians. In comparison, our study speaks mainly to over-the-counter healthcare products, which consumers purchase directly and rely on public healthcare-related information to make purchase decisions. Also, Ching et al. (2016) include news articles and television news program information, but they ignore the existence of user-generated content and TV shows like *The Dr. Oz Show*, which are believed to be highly influential in consumers' healthcare choices. Technically, Ching et al. (2016) employ humans for content coding of each news article, which is time-consuming and has little economy of scale. In contrast, we employ a scalable approach to content coding the textual information. Other related papers include Chintagunta et al. (2009) and Kalra et al. (2011). Chintagunta et al. (2009) focus on the learning process of doctors, and their data-coding design does not allow for rich textual features as covered by our study. Similarly, Kalra et al. (2011) study the impact of negative and positive media coverage on physicians' beliefs about the quality of a prescription-based diabetes drug. In sum, all three papers above study the effect of news coverage on prescription choices, which are made by doctors rather than patients. Further, they do not fully extract the content features of the textual information and do not consider relevant information from some other sources.

Our research is also related to the literature of the impact of user-generated content on demand (Chevalier & Mayzlin, 2006; Huang & Chen, 2006; Berger et al., 2010; Kalra et al., 2011, Liu et al., 2017, etc.). Basuroy et al. (2003) find that positive (negative) reviews increase (decrease) box office revenues. Berger et al. (2010) find that both positive and negative reviews can increase book sales. Most of the research used only the counts or simple features (e.g., length, sentiment) of the texts to study the effect on demand, due to the difficulty of processing textual data. However, the rich language features are ignored, which makes it impossible for us to understand what aspect of textual information is affecting consumers' choice.

More broadly speaking, this paper is related to the literature on applying natural language processing to marketing (e.g., Lee & Bradlow, 2011; Tirunillai & Tellis, 2012; Lee et al., 2017; Liu et al., 2017). Instead of solely relying on human coding or feature engineering, we also employ the state-of-the-art deep learning approach (LeCun et al., 2015) to content code all textual information from different sources.

Our paper also relates to the communication literature on media hype (e.g., Vasterman 2005), which describes the phenomenon of self-inflating media coverage on one specific story or topic. Vasterman (2005) uses a case study to examine the dynamics of media-hype. As a more specific research on hype in medical news, Zuckerman (2003) conducts three case studies of how companies shape news coverage of medical products. Goel et al. (2012) study the online information diffusion structures with seven examples. They find that most trends on social media form only after mentions from a few dominant individuals or news outlet amplification. In contrast, our paper presents empirical evidence of media hype on healthcare products—*The Dr. Oz Show* serving as the key event and news articles forming consonant news waves.

2.3 Data

We combine several datasets, including product data from Amazon, multi-media data from *The Dr. Oz Show*, research articles from academic journals, newspaper articles, magazine articles, and so on. This section explains each set of data sequentially.

2.3.1 Amazon Product Data

Table 1. Product-level Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
Number of Reviews	7.664	26.439	1	847
Price	23.704	24.149	0.010	1025.380
Average Rating	3.490	1.353	1	5
Average Ranking	104765.500	116826	55	1205512
Advertisement Spending (000)	6.243	111.648	0	4771.600

The first dataset we use is Amazon product data from 1996 to 2014,[10] including product reviews and product-level metadata. We focus on the category of health and personal care and further select out the weight loss subcategory. We replace the cross-sectional data on price with the time-series price information for each product, enabled by the API developed by Keepa.com. For each product, we collect the monthly advertisement spending data from Ad$pender database at the brand level[11]. In total, there are 6007 weight loss products and 160,000 consumer-generated reviews. Table 1 shows the summary statistics at the product level. Notice that the ranking is measured in the entire health and personal care category. Table 2 shows some examples of weight loss products in our sample. Unfortunately, we do not observe the exact sales of each product from Amazon.com. The same data limitation is addressed in Mayzlin and Chevalier (2006) by using sales rank to proxy sales, supported by a linear relationship between ln(sales) and ln(sale rank). We decide not to use this approach for two reasons. First, Amazon.com calculates sales rank based on the full history of sales rather than the sales per period.[12] Therefore, the sales rank data does not meet our need to measure sales in each period. Second, the data period (pre-2015) does not allow a time-series observation of sales rank. Due to these reasons, we instead use the number of customer reviews for each product in each time period as a proxy for sales. This choice is motivated by the observation that number of reviews has been widely used as a proxy measure for demand (bookings) estimate of Airbnb properties.[13] This measure is admittedly imperfect because of selection issues about when and

[10] Made available by Julian McAuley, UCSD. http://cseweb.ucsd.edu/~jmcauley/

[11] The user manual is available here: http://products.kantarmediana.com/documents/AdSpenderManual.pdf

Chapter 1 [12] For details of Amazon best sellers rank, please see https://www.amazon.com/gp/help/customer/display.html?nodeId=525376

[13] Please see http://insideairbnb.com/about.html for details. Please see appendix E for a further comparison between the two proxy measures, which further supports the usage of count of reviews as the proxy measure for sales.

who writes reviews.[14] However, it is the best available measure of healthcare product demand for researchers.[15] We acknowledge this is a limitation of the paper.

2.3.2 Products Included in Analysis

From the products' titles, we manually extract the key ingredients. For example, the ingredient in the first product title from Table 2 is garcinia cambogia, whereas that in the second title is raspberry ketones. For each ingredient, we check whether it has been recommended on *The Dr. Oz Show*. Among all the products, 2186 contain the ingredients mentioned by Dr. Oz. Below we list all ingredients that are identified from the product's titles.

- **Ingredients mentioned by Dr. Oz**: Garcinia cambogia, green coffee bean, raspberry ketone, saffron extract, forskolin, Conjugated Linoleic Acid (CLA), safflower oil, Yakon syrup, buckthorn, moringa, mulberry, coconut oil, hoodia, acai, bitter orange, apple cider vinegar, glucomannan, chitosan, 7-keto.
- **Ingredients not mentioned by Dr. Oz**: Caralluma, citrimax, Pina Colada, sesamin.

In addition to the products that contain the ingredients listed above, some products contain no clear ingredients in their titles (e.g., the last title in Table 2), while some do not work in a similar way as the "Oz products."[16] Textual information of these other products is not collected.[17]

Table 2. Examples of Product Titles

1	Garcinia Cambogia Extract by NewLife Botanicals
2	NatureWise Raspberry Ketones Plus+ Weight Loss Supplement and Appetite Suppressant
3	Lipozene Diet Pills - Maximum Strength Fat Loss Formula - 1500mg - 30 Capsules
4	Power Pops-hoodia Weightloss Lollipops-30ct Variety Pack
5	nuYou Labs Green Coffee Bean Extract with GCA Chlorogenic Acid - Highly Effective Natural Weight Loss Diet Supplement
6	One XS Weight Loss Pills (X-Strength) Prescription Grade Diet Pill. No Prescription Needed. Fast Proven Results. Weight Loss Guarantee

[14] It is a known fact that not every consumer writes reviews. Given that we focus on a single product category (weight loss), it is not unreasonable to assume that the fraction of consumers who write reviews is relatively similar across products. Therefore, the demand measure can be a multiple of the number of reviews and the multiplier is the same across products. This assumption implies that the parameter estimates in section 4.4 are unaffected by this measurement choice.

[15] We also explored the Nielsen scanner data. However, the weight loss products covered in Nielsen data are scarce and have little overlap with those mentioned by Dr. Oz. Moreover, other third-party datasets contain only ingredient-season level sales information.

[16] The active ingredients in the "Oz products" are all natural products. However, other products in the weight-loss category are made of artificial chemical compounds or use physical mechanisms to assist weight loss.

[17] Other non-Oz products (not good comparisons because of the mechanism) that are treated as outside options include Diuretic, Diurex, Enema, Hydroxycut, Nuphedrine, and Ornithine.

7	Molecular Research Labs Diet Supplement, Garcinia Cambogia Extract, 750 mg, 60 Count
8	NOW Foods Liver Detoxifier and Regenerator, 90 Capsules
9	Eden Pond Ketones Liquid Diet Drops Best Fat Burner Weight Loss That Works, Raspberry, 2 Fluid Ounce
10	Trimspa x32 Rapid Release Weight Loss 70 Capsules

2.3.3 Multi-media Data and Articles

For all the ingredients identified from the Amazon product titles, we collect all the publicly available information, including the script of *The Dr. Oz Show*, research articles from academic journals, and news articles. We collect research articles and news articles from the ProQuest Central database. ProQuest Central brings together many of the most used databases to create the most comprehensive, diverse, and relevant multidisciplinary research database available. It provides access to databases across all major subject areas, including business, health and medical, social sciences, arts and humanities, education, science and technology, and religion. The collection includes thousands of full-text scholarly journals, newspapers, magazines, dissertations, working papers, market reports, audio and video works, blogs, podcasts, websites, books, and trade journals. In our study, we define news article as all those included in ProQuest Central except for peer-reviewed research articles (e.g., newspaper, magazines, blogs, etc.). In Table 3, we list some examples of peer-reviewed journals from which we collect the related research articles. Please see the appendix B for the full list of journals.

Table 3. Examples of Peer-reviewed Journals for Research Articles

Journal of Food Science and Technology
Pharmaceutical Research
The British Journal of Nutrition
Journal of Dairy Science
Scientific Reports (Nature Publisher Group)
BioMed Research International
Marine Drugs
Obesity Research
International Journal of Molecular Sciences
Journal of Clinical Pharmacy and Therapeutics
Nutrition
Journal of Diabetes & Metabolic Disorders

When collecting the textual information, we include only articles with the keyword contained in the document title, in case the whole article is about something else but only mentioned the keywords in a trivial way. We find that all ingredients have related research findings except for Yakon syrup, apple cider vinegar, mulberry, Pina Colada, and citrimax.[18] Every ingredient has related news article coverage. To make it easier to compare the information intensity (frequency), we categorize the ingredients into two categories based on whether they were recommended by Dr. Oz. Table 4 displays the number of articles/shows/reviews covering the ingredients included in our analysis, by category. In the table, the "Oz category" includes all ingredients recommended by Dr. Oz and vice versa for the "non-Oz category". Note that the count of reviews for each ingredient measures the total number of reviews for all products that contain the ingredient under consideration. On an average, the number of news articles covering the ingredients in the "Oz category" is almost three times that of ingredients in the "non-Oz category". Moreover, the number of research articles that study the Oz-recommended ingredients is about six times that examining the non-Oz ingredients.

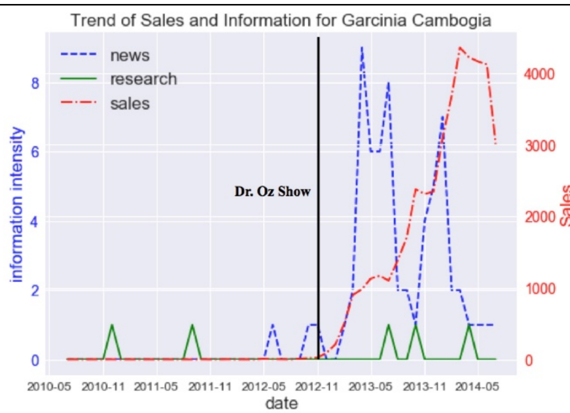
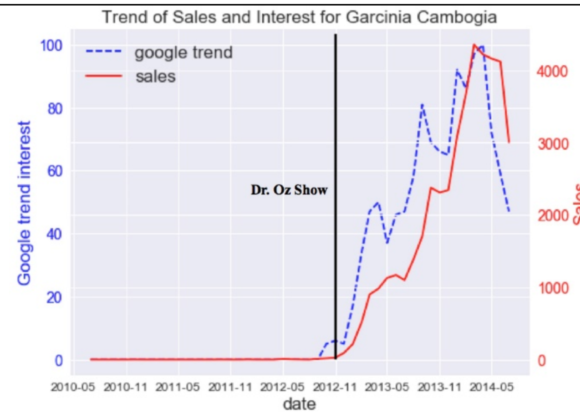
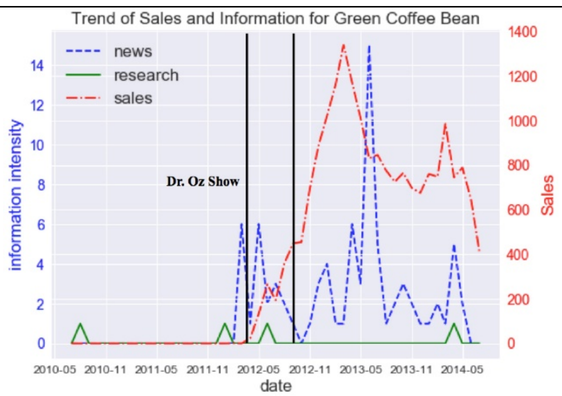
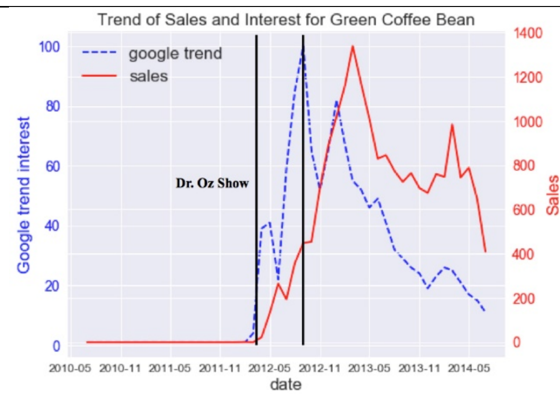
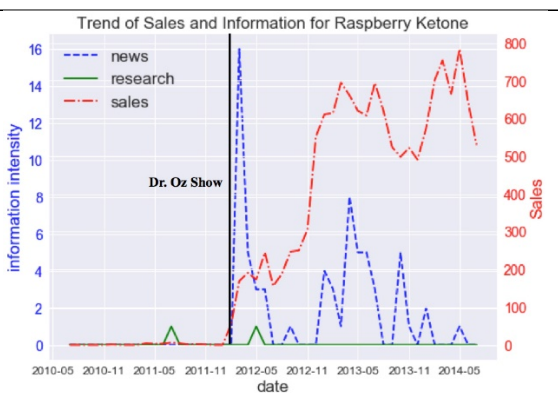
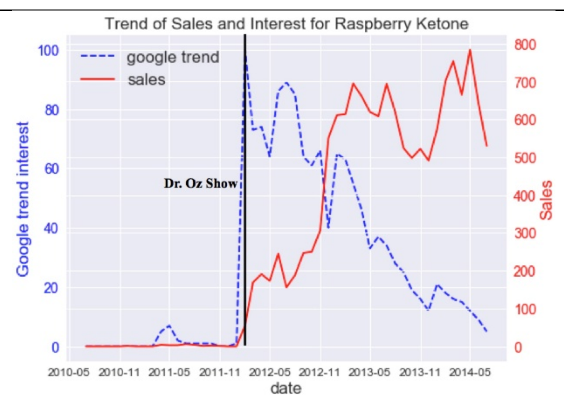
Table 4. Number of Shows, Articles, and Reviews

	Oz category	non-Oz category
average #(the Dr. Oz's Show) per ingredient	1.11	0
average #(news articles) per ingredient	68.28	22
average #(research articles) per ingredient	18.28	3.4
average #(reviews) per ingredient	4711.06	68.2

In Figure 1, we show time trends of sales, information intensity, and Google interest for products that contain a representative ingredient recommended on *The Dr. Oz Show*. Information intensity measures the number of text documents published in each time period (i.e., month in our analysis). According to the first column of figures, *The Dr. Oz Show* seems to have been playing a crucial role in increasing the sales of the products as well as the intensity of information that contain the recommended ingredients. According to the second column of figures, the discrepancy between Google trend and sales trend implies that there is some time lag between searching and purchasing. Though both actions increase after *The Dr. Oz Show*, the increase of searching interest happens more promptly. The data evidence in Figure 1 indicates the possibility that consumers' purchase decisions may be largely affected by the information from different sources after *The Dr. Oz Show* rather than the show alone.

[18] Pina Colada comprised several tropical fruits. This can be one reason why there is no research article about it. Because the sellers do not separately list the fruits in the product titles, we directly use "Pina Colada" to capture this type of products.

Figure 1. Time Trends

Product (Date on Show)	Trend of sales, news and research articles	Sales trend and Google trend
Garcinia Cambogia (Nov 5, 2012)		
Green Coffee Bean (April, 2012) (Sep, 2012)		
Raspberry Ketone (Feb, 2012)		

Note: the vertical lines point to the time when Dr. Oz mentioned the ingredient in his show.

The statistics shown in Figure 1 and Table 4 imply that *The Dr. Oz Show* may be correlated with the increasing sales of the products containing the recommended ingredients. However, the increasing sales may not be caused by the show alone, given the abundance of information from other sources. For the rest of the paper, we employ natural language processing techniques and econometric models to analyze the effect of different sources of textual information on consumers' choices.

2.4 Methodology

To process the unstructured textual data introduced in Section 3, we go through the following procedure:

1. Use a CNN model to identify eight content features.

2. Extract four language features using traditional NLP.

3. Include content features and language features in the dynamic panel linear model, and analyze the effect of each feature on sales.

The first two steps aim to extract all textual features, detailed in Section 4.1. The models used to extract these features are introduced in Section 4.2 and 4.3. Specifically, Section 4.2 describes the CNN model (step 1) while Section 4.3 presents the traditional NLP models (step 2). Finally, section 4.4 specifies the econometrics model that estimates the impact of textural features on demand (step 3).

2.4.1 Textual Features to Extract

In total, we extract thirteen textual features. Please see Table 5 for the summary statistics of all textual features.

For the first eight textual features listed in Table 5, we use CNN to extract them from each information source. Before describing CNN in Section 4.2, we first explain these eight features. We believe that both informative and persuasive information in textual documents affect consumer decision making. Therefore, based on the work of Resnik and Stern (1977) and Lee et al. (2017), we choose features that can represent the persuasive aspects and/or informative aspects of the content information. The persuasive aspects include the use of positive emotion (denoted by emotionP in Table 5), negative emotion (denoted by emotionN) and humor (denoted by humor), because emotional and humorous content are identified as drivers of virality (e.g., Porter & Golan, 2006; Berger & Milkman, 2012; Berger, 2014). The informative aspects include mention of price (or price comparison), mention of product/brand name (denoted by pnames), side effects (denoted by side), and performance of the product (denoted by performN and performP, see Resnik & Stern, 1977 and Lee et al., 2017 for the definition of performance). To extract these features, CNN is used for the following two reasons. First of all, CNN has outstanding performance on NLP tasks (e.g., Liu et al., 2017; Timoshenko & Houser, 2017), and it fits our need as a scalable supervised prediction technique to detect whether a text document contains a specific feature. Second, there are no well-established traditional NLP tools to content code these features.

In addition to these eight features, four other language features, i.e., complexity, sentiment, subjectivity, and ambiguity, are extracted. We choose these four features because prior literature found that they have a significant effect on product sales or consumer decisions (e.g., Berger et al., 2010; Ghose & Ipeirotis, 2011; Gong et al., 2017). We use traditional NLP tools to extract these four features because they are supported by large-size external corpus and have shown good performance and robustness in extracting these four features. We explain each of the feature extraction tasks in detail in Section 4.3.

Table 5. Description of Textual Features

Variable[1]	Method[2]	Description
pname_s	DL	dummy, equals 1 if product/brand name is mentioned in the text from source s in each period.
side_s	DL	dummy, equals 1 if side effect is mentioned in the text from source s in each period.
price_s	DL	dummy, equals 1 if price information is mentioned in the text from source s in each period.
humor_s	DL	dummy, equals 1 if humor is used in the text from source s in each period.
performP_s	DL	dummy, equals 1 if the product is considered as effective in the text from source s in each period.
performN_s	DL	dummy, equals 1 if the product is **not** considered as effective in the text from source s in each period.
emotionP_s	DL	dummy, equals 1 if positive emotion appears in the text from source s in each period.
emotionN_s	DL	dummy, equals 1 if negative emotion appears in the text from source s in each period.
complexity_s	Trad.NLP	the measure of difficulty to understand the focal textual information from source s in each period.
sent_s	Trad.NLP	the measure of sentiment of the focal textual information from source s in each period.
subj_s	Trad.NLP	the measure of subjectivity of the focal textual information from source s in each period.
ambiguity_s	Trad.NLP	the measure of how ambiguous of the focal textual information from source s in each period.
intensity_s	Summation	the measure of how many units of the textual information from source s in each period.

Note: 1. $s \in \{Oz, res, news, rew\}$, where “rew” refers to customer review, and “res” refers to “peer-reviewed research articles”. 2. The method “DL” means deep learning, specifically CNN; “Trad.NLP” means tranditional NLP methods.

The last feature we consider is intensity. Now we describe how we define and construct the intensity measure. As explained in Table 5, intensity is a measure of the density of the textual information in each period. We measure intensity with a simple approach by counting the frequency of all existing pieces of information at each time point. For example, the intensity for "research" of ingredient j in period t would be 2 if two new research articles about j were published in period t. The intensity defined in this way serves our intention to measure the marginal effect of an additional article/show/review on product sales. For each information source, we use "cumulative sum" for the intensity measure and "average" for other features,[19] because "sum" captures the intensity of the information, while "average" provides an overall measure for all other language features.

2.4.2 Informational Content Extraction with CNN

As mentioned before, we use CNN to extract the first eight features listed in Table 5. We follow two steps to label each piece of textual information, which could be a product review, a newspaper article, a research article, or a script of an episode of *The Dr. Oz Show*. First, we hire workers through AMT and tag 3,000 messages for a variety of textual contents. Subsequently, using the labeled contents, we train a CNN model to content code the full set of messages (more than 160,000 messages). Our CNN consists of four layers, as shown in Figure 2 (e.g., Kim, 2014; Liu et al., 2017; Timoshenko & Hauser, 2017). We briefly describe each layer of the CNN as follows.

- Layer 1: Word embedding.

The first layer is the word embedding or word vectors. Following the popular method to improve performance without a large supervised training set, we initialize word vectors with those obtained from an unsupervised neural language model. That is, the publicly available word2vec vectors trained on 100 billion words from Google News. Each vector has a dimension of 300 and was trained using the continuous bag-of-words architecture (Mikolov et al., 2013). New words are randomly initialized. As displayed in Figure 2, the first layer is the representation of the sentence, with each word represented by a 300-dimensional vector. With n being the total number of words in the text, the representation matrix is of dimension $n \times 300$. The i-th word is denoted as $\mathbf{v_i}$.

- Layer 2: Convolutional layer.

The convolutional layer applies convolutional operations with varying filters to the sentence representation in the first layer. The filter can be denoted as a vector $\mathbf{w} \in R^{1\times dh}$, which

[19] Both the "sum" and "average" for each language feature in period t are calculated across all the text documents published within period t. We also conducted robustness checks by considering not only period t, but also the past n periods, where $n \in \{1,2,3,4,5\}$. In other words, the "sum" and "average" are calculated across all the text documents published within period $t, t-1, \ldots, t-n$. The results remain largely consistent.

corresponds to a concatenation of all rows in a matrix from the second layer as shown in Figure 2. h is the size of the filter, and d is the dimension of the word embedding (i.e., 300). In our model, three different filter sizes are implemented (e.g., 3, 4, 5). The feature map for each filter with size h is a vector of the outputs of the convolutional operation, that is,

$$\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}]$$
$$c_i = \sigma(\mathbf{w} \cdot \mathbf{v_{i:i+h-1}} + b)$$

where $\sigma(\cdot)$ is a non-linear activation function, and we used $\sigma(x) = max\{x, 0\}$ in the neural network. $\mathbf{w}$ is the vector of linear weights and b is the bias (i.e., intercept), both of which are to be estimated. $\mathbf{v_{i:i+h-1}}$ is a concatenation of the vectors representing words i to $i + h - 1$; therefore, it is of dimension $dh \times 1$.

Layer 3: Pooling layer.

The pooling layer aims to transform the feature maps to a lower-dimensional vector, so to get the most salient textual information. The output is specified as

$$\mathbf{p} = [p_1, p_2, \dots, p_{mk}]$$

$$p_j = max\{c_1, c_2, \dots, c_{n-h+1}\}$$

where k is the total number of filters, and p_j corresponds to the output resulting from the filter of size h. We use 128 filters for each filter size h (h =3, 4, 5 in our network architecture). Therefore, there are in total $k = 128 \times 3 = 384$ filters.

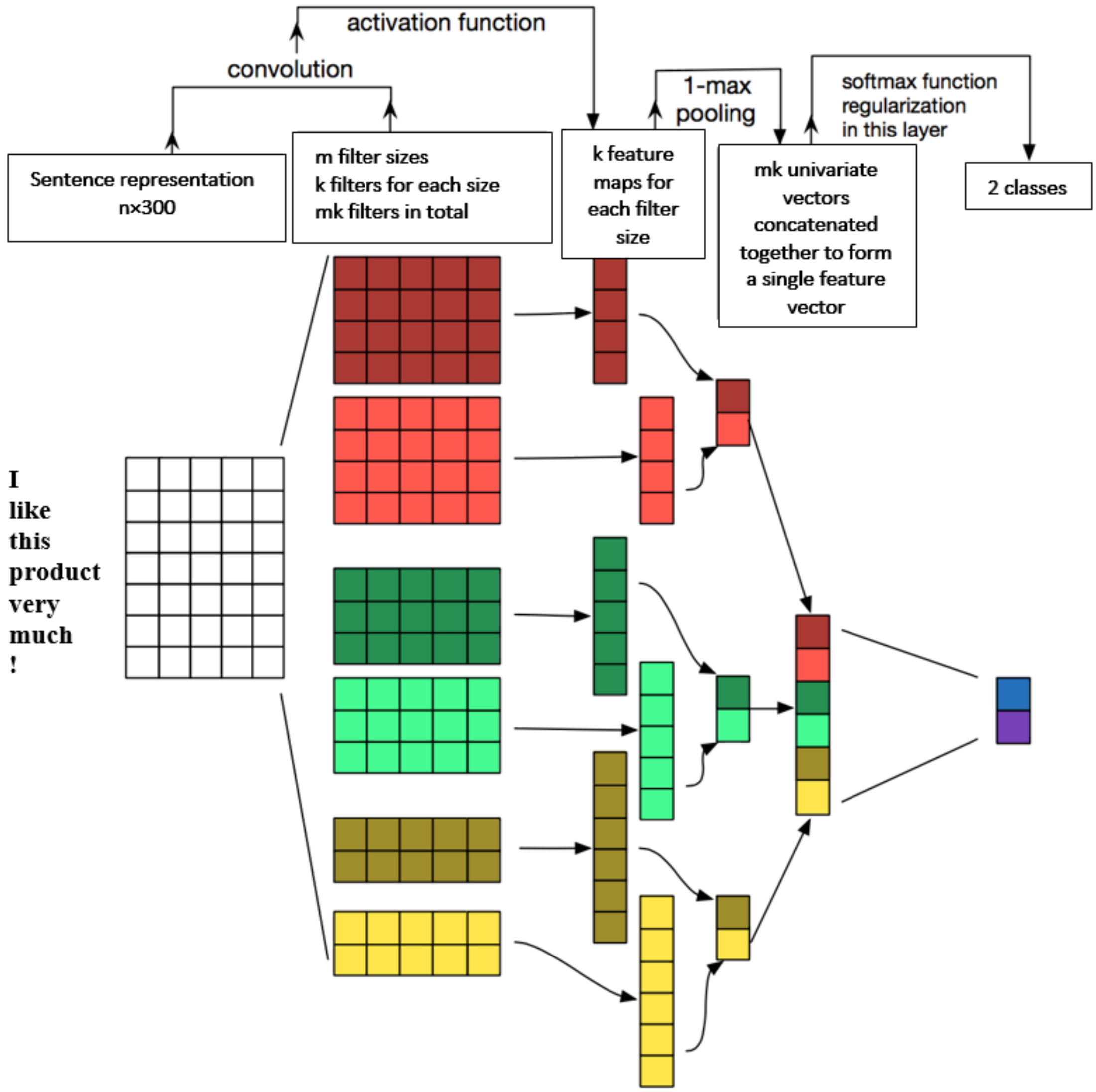
Layer 4: Softmax layer

The last layer in CNN is the softmax layer. This final layer takes the output of the pooling layer (i.e., $\mathbf{p}$) as input, and outputs the probabilistic prediction of whether a feature is contained in this text. Therefore, the output is a binary result y, which equals 1 if the text is classified as containing the feature under examination. The softmax specification is

$$y = softmax(\mathbf{W} \cdot \mathbf{p} + b)$$

where the weights $\mathbf{W}$ and bias b are to be calibrated through training the CNN.

Figure 2. Architecture of Convolutional Neural Networks for Sentence Classification



Note: in this figure, k=2, m=3, and the three filter sizes are 2, 3, 4; in our application, k=128, m=3, the filter sizes are 3, 4, 5. The figure is adapted from Figure 1 of “A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification,” by Zhang and Wallace (2016).

We use a mini-batch of size 64 while training the CNN. Following the rule-of-thumb, we set the drop-out rate as 0.5 in order to help prevent overfitting. We employed 10-fold cross validation to train the CNN. Three common criteria (i.e., precision, recall, accuracy) are used to evaluate the performance (Lee & Bradlow, 2011).[20] The CNN classifier's performance on a test sample of 1000 observations is shown in Table 6.

[20] $Precision = \frac{true\ positive}{true\ positive+false\ positive}$, $Recall = \frac{true\ positive}{true\ positive+false\ negative}$, $Accuracy = \frac{true\ prediction}{test\ sample\ size}$.

Figure 3 shows some examples of the classified text for each content feature.

Table 6. Performance on Content Coding

	pname	side effect	price info	humor	performP	performN	emotionP	emotionN
Precision (%)	89.8	88.9	77.6	85.7	88.5	82.6	72.3	80.8
Recall (%)	57.5	60.4	57.8	58.1	90.8	75.3	58.2	66.7
Accuracy (%)	85.7	94.2	94.7	96.9	86.8	89.5	79.0	85.9

Figure 3. Classified Text for Eight Content Features

Feature	Example Text Classified as Containing the Feature
pname	Eden Pond Labs Garcinia Cambogia has lead to me having significant weight loss.
side effect	I had awful gas and terrible stomach pain. The experience was horrible to say the least.
price info	Best price around by $20! Compared numerous websites and you were the best price by far!
humor	used these for yrs, not sure it works but still buying..lol.guess it must work since I live pasta and remain thin.
performP	I'm amazed!!! Works!! Really works!! I lost 11 pound in just one week and belly fat!!
performN	Did not perform as advertised, great waste of money. zero results.
emotionP	Great Product!!! If you have not tried this yet, then try it today. So glad I found it! thanks!
emotionN	I have tried several kinds of this pill and none of them has done anything for me. I am very disappointed.

2.4.3 Language Features Extraction Leveraging NLP Model

In addition to the eight content features extracted using CNN, we use traditional NLP techniques to extract four other language features widely documented in the literature.

- Complexity

To measure how difficult it is to understand the information, we consider two candidates. The first measure of text complexity we use is "sentence structure complexity" based on the syntactical structure, using the average height of the whole parse tree for each sentence. An alternative measure is a widely used metric called the Simple Measure of Gobbledygook index (SMOG) (e.g., Ghose & Ipeirotis, 2011). It measures the level of education needed to understand a piece of writing. A higher score means it is more difficult to comprehend the text. A 2010 study published in the Journal of the Royal College of Physicians of Edinburgh stated that "SMOG should be the preferred measure of readability when evaluating consumer-oriented healthcare material."

- Ambiguity

Following Gong et al. (2017), we use entropy over topics as a measure of ambiguity. The topics are discovered by using the Topic Modeling approach, namely, the Latent Dirichlet Allocation model (Blei et al., 2003).[21] When a text document contains a more balanced information for all aspects of the product, it indicates higher ambiguity. For example, a product review describing both the side effect and the taste of the product is more ambiguous than a review describing mostly the side effect but little taste information.

- Sentiment

We use the library TextBlob for sentiment analysis. TextBlob is a high-level library built on top of the NLTK library[22] (D'Andrea et al., 2015). As we pass text to create a TextBlob object, the TextBlob library performs the following processing over text: tokenize the text, i.e., split words from body of text; remove stopwords from the tokens; POS (part of speech) tagging of the tokens and select only significant features/tokens like adjectives, adverbs, etc.; pass the tokens to a sentiment classifier which classifies the text sentiment as positive, negative, or neutral by assigning it a polarity score between -1 to 1.

- Subjectivity

The subjectivity estimation came from computational linguistics (i.e., Pang & Lee, 2004). Intuitively speaking, subjective information expresses a very personal rather than a factual evaluation of the product. We follow the approach of Ghose and Ipeirotis (2011) to train the classifier. After classifying each sentence of a text document, we calculate the ratio of sentences classified as subjective and use the ratio as the subjectivity measure of the entire document.

The following Table 7 summarizes all textual features we have extracted from each information source. A set of observations can be made from Table 7. First, for some information sources, certain features are absent or have no variation. For example, *The Dr. Oz Show* never mentions the price information of any product under examination. Also, there is no humor used in any research article. These variables will not be included in our main analysis in step 5. Second,

[21] When there is no information, we give 2.302585, which is the entropy for uniform distribution.

[22] For the latest version, please see https://pypi.python.org/pypi/textblob

each language feature varies across different information sources. By comparing the mean value of each feature, we can get the following results.

(a) Subjectivity: Review>Oz>News>Research.

This is in expectation because the research articles are believed to be written based on fairly objective evidence, whereas reviews are written by consumers according to their personal experience and feelings.

(b) Complexity: Research>News>Oz>Review.

The complexity of each information source again matches our expectation. Research articles are the most difficult to understand and require the highest education level. News articles are written in an organized and formal way, therefore are harder to understand than the other two casual communication formats, i.e., talk show and customer word of mouth. Dr. Oz uses many medically sophisticated words, which makes the show a little bit harder to understand than consumer reviews.

(c) Sentiment: Review>Oz>News>Research.

On an average, customer reviews contain the most positive messages, while research is almost neutral.

(d) Ambiguity: Review>Oz>News>Research.

Consumers tend to mention many different aspects of the product, such as packaging, delivery, personal expectation, etc. On the other hand, research articles focus on one single important issue, i.e., the effectiveness of the focal ingredient.

(e) Contain negative emotion (emotionN): Review>News>Research>Oz.

Consumers are more likely to express emotions after experiencing a bad product or services. The objective of Dr. Oz Show is to recommend a certain ingredient. As he put it during the congressional hearing: “My job, I feel, on the show is to be a cheerleader for the audience.” Therefore, a negative emotion is rarely used in his language.

(f) Contain positive emotion (emotionP): Oz>Review>News>Research.

On his show, Dr. Oz tends to use exaggerating words and expressions to describe the products, unquestionably showing positive emotion. In comparison, research articles are evidence-based and written in an emotionless way.

(g) Mention the ineffectiveness (performN): Review>Research>News>Oz.

Again, Dr. Oz is focused mainly on the effectiveness of a certain ingredient on his show. However, consumers tend to share their disappointment through reviews when a product does not work for them.

(h) Mention the effectiveness (performP): Research>Oz>News>Review.

Research is surprisingly similar to The Dr. Oz Show, regarding the frequency of mentioning the effectiveness of the focal ingredient. Consistent with (d), customer reviews may not focus on the effectiveness but on other aspects of the product.

(i) Mention price related information (price info): News>Review>Research>Oz.

Dr. Oz never mentions price, in order to detach from any specific merchant. In contrast, news articles and customer reviews have less of this concern, thus describe many attributes of a product (e.g., price) in addition to its efficacy.

(j) Mention side effect: Review>News>Oz=Research.

The "side effect" feature is assigned value 1 if the text document indicates that the product does have a side effect. The value is 0 if there is no mention of side effects, or if the textual information covers some discussion on side effects but implies that none are confirmed or experienced. Consumers have the most to say when they experienced a side effect. But The Dr. Oz Show never points out a product's side effects when recommending the ingredient.

(k) Use humor: Humor is used only in customer reviews.[23] As a persuasive feature, it is well adopted by first hand users of the product.

(l) Mention product name: News>Research>Oz>Review.

A customer comments within the product's web page and therefore does not need to mention product names. But news articles and research articles have to make the focus of the whole article—the product/ingredient—salient in front of readers' eyes.

2.4.4 Econometric Model

Our analysis follows the Arellano and Bond (1991)'s approach (henceforth AB). Given the panel nature of our dataset, the AB approach can account for the autoregressive dynamics and potential endogeneity issues.

The model can be characterized as follows:

$$y_{jt} \quad = \alpha_0 + \boldsymbol{\alpha_1}\mathbf{T_{jt}^{Oz}} + \boldsymbol{\alpha_2}\mathbf{T_{jt}^{research}} + \boldsymbol{\alpha_3}\mathbf{T_{jt}^{new}} + \boldsymbol{\alpha_4}\mathbf{T_{jt}^{review}} + \beta_1 y_{j,t-1} + \beta_2 \mathbf{x_{jt}} + \xi_j + \epsilon_{jt} \quad (1)$$

where y_{jt} is the sales of product j in period t. ϵ_{jt} is the unobserved product-time specific idiosyncratic demand shock, which captures the information from hospital visits, friends and families. ξ_j captures the unobserved product characteristics. We include the sales in the

[23] Though Dr. Oz sometimes tells jokes on his show, the trained CNN did not detect humor from the shows where he recommended the ingredients included in our sample.

Table 7. Summary Statistics of Language Features

News (Count:1339)

	subj	complexity	sent	amb.	emN	emP	perN	perP	price	side	humor	pname
Mean	-0.395	11.535	0.113	0.504	0.037	0.276	0.025	0.767	0.076	0.010	0	0.849
Std	0.360	2.048	0.092	0.377	0.188	0.447	0.155	0.423	0.265	0.102	0	0.357
min	-1	4.714	-0.433	0.055	0	0	0	0	0	0	0	0
50%	-0.391	11.507	0.109	0.533	0	0	0	1	0	0	0	1
max	1	29.924	0.600	2.019	1	1	1	1	1	1	0	1

Research (Count:346)

	subj	complexity	sent	amb.	emN	emP	perN	perP	price	side	humor	pname
Mean	-0.649	13.590	0.066	0.360	0.006	0.072	0.029	0.853	0.003	0	0	0.795
Std	0.334	2.261	0.093	0.357	0.076	0.259	0.168	0.355	0.054	0	0	0.404
min	-1	7.679	-0.250	0.079	0	0	0	0	0	0	0	0
50%	-0.745	13.663	0.064	0.130	0	0	0	1	0	0	0	1
max	0.714	21.641	0.550	1.576	1	1	1	1	1	0	0	1

Oz (Count:20)[24]

	subj	complexity	sent	amb.	emN	emP	perN	perP	price	side	humor	pname
Mean	-0.188	7.154	0.163	0.578	0	0.7	0	0.85	0	0	0	0.55
Std	0.216	1.068	0.080	0.442	0	0.470	0	0.366	0	0	0	0.510
min	-0.571	5.684	0.006	0.082	0	0	0	0	0	0	0	0
50%	-0.226	6.945	0.163	0.592	0	1	0	1	0	0	0	1
max	0.294	9.726	0.348	1.516	0	1	0	1	0	0	0	1

Review (Count:150744)

	subj	complexity	sent	amb.	emN	emP	perN	perP	price	side	humor	pname
Mean	0.034	6.821	0.166	0.987	0.080	0.252	0.241	0.637	0.052	0.029	0.001	0.176
Std	0.521	2.016	0.214	0.415	0.272	0.434	0.427	0.481	0.222	0.167	0.017	0.381
min	-1	3.1291	-1	0.066	0	0	0	0	0	0	0	0
50%	0	6.950	0.16	0.977	0	0	0	1	0	0	0	0
max	1	21.194	1	2.996	1	1	1	1	1	1	1	1

Note: amb. , emN, emP, perN, perP means ambiguity, emotionN, emotionP, performN, performP accordingly.

[24] Even though there are only 20 Oz shows, the recommended ingredients apply to many products. In our full sample of weight loss products, there are 36% products in our sample containing the ingredients mentioned by Oz shows.

previous period, $y_{j,t-1}$, in the model to capture the effect from returning customers, if there is any.[25] $\mathbf{x_{jt}}$ includes a group of product characteristics: average rating for product j up to period t (avg_rate), price of product j in period t (price), and normalized advertisement spending for the product's brand (n_ads). $\mathbf{T_{jt}^s}$ are the textual features of the four information sources, where $s \in \{Oz, research, news, review\}$ denoting *The Dr. Oz Show*, research articles, news articles, and reviews. Each $\mathbf{T_{jt}^s}$ contains thirteen features:

$$\mathbf{T_{jt}^s} = \begin{Bmatrix} sent\_s_{jt}, complexity\_s_{jt}, ambiguity\_s_{jt}, subj\_s_{jt}, publicity\_s_{jt}, price\_s_{jt}, side\_s_{jt}, \\ hum\_s_{jt}, pname\_s_{jt}, emotionN\_s_{jt}, emotionP\_s_{jt}, performN\_s_{jt}, performP\_s_{jt} \end{Bmatrix}$$

For ingredients that are not mentioned in a period or information source, the text feature values are set to zero for the binary variables (i.e., $publicity_{s_{jt}}, price_{s_{jt}}, side_{s_{jt}}, hum\_s_{jt}$, $pname\_s_{jt}, emotionN\_s_{jt}, emotionP\_s_{jt}, performN\_s_{jt}, performP\_s_{jt}.$), meaning there is no mention of these content features. For sentiment, we set it to zero, which represents neutrality when there is no information. Similarly, subjectivity is set to zero, indicating there is neither subjective opinions nor objective facts. Complexity is zero when there is no information, and ambiguity is equal to the entropy over uniform posterior distribution over topics.[26]

2.4.5 Endogeneity

All the textural features $\mathbf{T_{jt}^s}$ as well as the price p_{jt} can be endogenous because: first, they may be correlated with the unobserved product-level characteristics ξ_j; second, they may be correlated with current and possibly past realizations of the errors ϵ_{jt}. Arellano and Bond (1991) accounts for the endogeneity issues by taking all potential orthogonality conditions into account and applying a Generalized Method of Moments (GMM) estimation. Following the AB approach, we first remove the unobserved individual effect by first-differencing equation (1):

$$\Delta y_{jt} \quad = \boldsymbol{\alpha_1} \boldsymbol{\Delta} \mathbf{T_{jt}^{Oz}} + \boldsymbol{\alpha_2} \boldsymbol{\Delta} \mathbf{T_{jt}^{research}} + \boldsymbol{\alpha_3} \boldsymbol{\Delta} \mathbf{T_{jt}^{new}} + \boldsymbol{\alpha_4} \boldsymbol{\Delta} \mathbf{T_{jt}^{review}} + \beta_1 \Delta y_{j,t-1} + \beta_2 \boldsymbol{\Delta} \mathbf{x_{jt}} + \Delta \epsilon_{jt} \quad (2)$$

This transformation removes both the constant term and the product fixed effect, so the first endogeneity issue is resolved. We then use the lagged values (up to the second lag) of the endogenous variables (i.e., $\mathbf{T_{jt}^s}$ and p_{jt}) as instruments to account for the second endogeneity issue, caused by the potential correlation between the differenced lagged dependent variable and the disturbance process. The lags are valid instruments because they are correlated with the first differences of the endogenous variables but uncorrelated with $\Delta\epsilon_{jt}$. We use lagged

[25] By randomly checking the product reviews that mentioned the time lag between purchase and writing reviews, we find that more that 90% lags are less than one month. This is why we use the number of reviews in the current month to measure sales.

[26] We conduct multi-collinearity tests to make sure all variables included in the estimation are immune to the multi-collinearity problem.

values up to the second lag because there is no presence of second-order serial correlation in the error terms, according to the Arellano-Bond test for AR(2) in Table 8.[27]

2.5 Results

2.5.1 Model Specification Tests

To test the validity of the instruments, we use Hansen test of over-identification. As reported in Table 8, the over-identification restrictions are valid, namely, the instruments are jointly valid. Moreover, the Arellano-Bond tests for AR(1) and AR(2) reported in the table indicate that errors are not second-order serially correlated.[28] The first order serial correlation is expected due to the lagged dependent term.

2.5.2 Main Findings

The estimation results for equation (1) are summarized in Table 8. The table reports the results for both OLS (first column) and AB approach (second column). Observations are at the product-month level. Features that have no variations (e.g., price_res) are not included in the model. Please see the correlation matrix of all the variables in Appendix D. We discuss the results for each information source one by one.

1). The Dr. Oz Show. Across all textual features, *The Dr. Oz Show* has no significant effect on the product's sales except for the negative effect of the language complexity. This stands against the figures we show in Figure 1 that imply a huge boost in sales of the recommended products. The finding indicates that the increasing sales of the focal product happened together with *The Dr. Oz Show* but were not directly driven by it. We further explore this puzzle in section 5.3.

2). News articles. Given other factors controlled, the intensity of news articles has a significant positive effect on the product's sales. When there are more news articles mentioning a product's name or its brand name, the product's sales also increases. Moreover, a product's sales increase when the news articles have higher sentiment measure, written in a more ambiguous way or with a more subjective tone. This set of findings implies that news articles, as a fairly common media channel, have significantly affected the general public's choices in the healthcare domain. It is noteworthy that news articles are not often written with evidence that can stand the test of time. The results indicate that consumers' purchasing decision may more likely be affected by news articles with subjective arguments rather than objective facts. Therefore, the content of healthcare-related news articles should be strictly supervised,

[27] Otherwise we need to use lags up to the third lag.

[28] We also considered including lags for text features, please see appendix C for the results. The implication is consistent with the main findings.

especially when they mention a particular product/brand name, or claim the product is effective.

3). Peer-reviewed research articles. The first prominent result about research articles is the positive effect of intensity. Our initial belief may be that research articles are mostly circulated in academia hence have little effect on consumer choice. Our results tell a fact that scientific research findings do affect consumer choices significantly. Though ordinary people usually do not read peer-reviewed research articles for medical advice, there may be several possible mechanisms for research articles to have a strong impact on sales. One is that due to high credibility, research articles are often ranked at top positions by search engines such as Google and Bing. When looking for healthcare products, consumers might encounter research articles on search engine result pages. Moreover, doctors often disseminate research findings to patients to educate patients. Furthermore, research article findings are widely shared on social networks such as Facebook or Twitter. Since these three channels are very popular, they could contribute to the effectiveness of research articles. In addition to intensity, from Table 8, we can also see that sales are negatively affected by the complexity and ambiguity of research articles, as well as emotions contained in the research articles. These results provide a strong incentive for researchers to write papers in a simpler way, so that their findings can have a higher impact in the non-academic world. Moreover, the negative effects from ambiguity and emotion imply that the readers of scientific research articles expect the contents to be neutral, serious and focus on some key aspect of the concerned product. Similar to news articles, a product's sales also increase when the research article mentions the product names.

4). Product reviews. For customer reviews, significant effects on sales come from cumulative average numerical rating and the sentiment of the review content. As expected, the numerical rating positively affects a product's sales.[29] However, given the same rating, the sentiment of a product review has a negative effect on sales. This may be caused by the substitution between positive words and neutral description of product characteristics which consumers value more. Another explanation could be that, when the rating is at a given level, the review's credibility gets questioned when the words used are too positive.[30]

---

[29] As we explained previously, the sales measure is constructed with the number of reviews, so the intensity measure for product reviews is not included as an explanatory variable.

[30] Associated Press (2015). *Fake Online Reviews: Here Are Some Tips for Detecting Them*. Retrieved from https://www.nbcnews.com/business/consumer/fake-online-reviews-here-are-some-tips-detecting-them-n447681

Table 8. Estimation Results with Text Features

VARIABLES	Model 1. OLS		Model 2. AB Estimates	
	Parameter	s.e.	Parameter	s.e.
Sales_lag	0.894***	0.00531	0.640***	0.0568
price	0.000708	0.00119	-0.00201	0.00869
avg_rate	0.379***	0.112	2.477*	1.429
n_ads	-0.0675	0.0969	0.518	2.026
ambiguity_oz	3.932***	0.448	-6.197	7.017
ambiguity_res	1.907***	0.645	-3.636*	1.940
ambiguity_news	3.326**	1.576	5.487***	1.697
ambiguity_rew	-0.369	0.245	-0.125	0.681
sent_res	5.144	6.539	10.13	12.13
sent_news	10.97	8.393	17.59**	7.010
sent_rew	-1.019***	0.221	-5.272***	2.019
subj_oz	-4.075**	1.639	-17.50	10.97
subj_res	2.176	2.084	2.132	1.994
subj_news	1.668**	0.587	3.299***	1.120
subj_rew	-0.276	0.198	-0.342	0.689
complexity_oz	-3.907***	0.290	-8.117***	2.988
complexity_res	-0.747***	0.176	-1.112***	0.346
complexity_news	-0.0930	0.101	-0.183	0.266
complexity_rew	-0.0667	0.0405	-0.194	0.190
intensity_oz	19.54***	1.551	11.71	18.25
intensity_res	1.491	1.625	2.910*	1.516
intensity_news	0.000689	0.136	0.0730**	0.025
price_news	-1.389	1.018	-1.449	1.441
price_rew	-0.226	0.308	1.135	1.389

Table 8. Estimation Results with Text Features (Continued)

VARIABLES	Model 1. OLS		Model 2. AB Estimates	
	Parameter	s.e.	Parameter	s.e.
emotionp_res	-2.405	1.803	-4.413**	1.974
emotionp_news	-0.843	0.838	-0.846	0.953
emotionp_rew	0.931***	0.263	1.600	1.104
emotion_news	-0.412	0.865	-0.230	1.899
emotionn_rew	-0.158	0.176	-0.415	1.287
performp_res	0.178	1.018	0.944	1.399
performp_news	0.152	0.783	-1.175	1.414
performp_rew	1.597**	0.635	1.092	1.221
performn_res	3.486**	1.447	2.706	1.891
performn_news	-6.451	5.492	-9.619	14.72
performn_rew	-0.889*	0.458	-0.769	0.990
pname_oz	7.422***	1.172	17.99	12.48
pname_res	3.462	2.121	4.320*	2.278
pname_news	7.038***	1.891	13.11***	3.957
pname_rew	1.083***	0.161	-0.00695	1.497
hum_rew	2.261	2.000	-23.40	24.75
side_news	-2.008	3.096	3.878	6.274
side_rew	0.455	0.287	-3.385	2.401
#Obs.	6688		4965	
Adjusted R^2	0.7873		0.7974	
Hansen Chi^2	-		713.28	p-value=1.000
Test for AR(1)	-		-4	p-value<0.001
Test for AR(2)	-		1.81	p-value=0.070

Note: #Obs. for Model 2 is smaller than that of Model 1 because of the first-differencing operation;
The null hypothesis of the Hansen test is Ho: over-identification restrictions are valid;
The null hypothesis of the Arellano-Bond test for serial correlation is Ho: no autocorrelation;
Robust standard errors, clustered at ingredient level;
***p<0.01, ** p<0.05, * p<0.1

To sum up, the results above help us understand how different information sources affect consumers' healthcare product choices. To better understand the mechanism, we measure consumers' perceived credibility of each information source and test whether credibility can rationalize the differentiated effects of various information sources. Specifically, we conduct a survey enabled by AMT and Qualtrics.com on how consumers perceive the credibility of the four sources of information. We directly asked consumers who access all four sources of information to rate the credibility level of each source from 1(least credible) to 7(most credible). The following Figure 5 and Table 11 show the survey results. Not surprisingly, research articles are considered as the most credible information source. Interestingly, Dr. Oz Show is considered the least credible one. The credibility score for news articles and customer reviews are intermediate and almost comparable. The survey results support our empirical findings that Dr. Oz's suggestion is less likely to be taken seriously by consumers, given his low credibility, which is a gratifying fact from a consumer welfare perspective.

Table 9. Estimation Results without Text Features

VARIABLES	AB Estimates	
	Parameter	s.e.
Sales_lag	0.742***	0.125
price	0.00370	0.00790
avg_rate	4.056***	0.730
n_ads	2.911	2.001
intensity_oz	1.233	3.029
intensity_res	0.188	0.777
intensity_news	-0.0733	0.136
#Obs.	4,965	
Adjusted R^2	0.4184	
Hansen Chi2	259.75	0.993
Test for AR(1)	-3.22	p-value=0.001
Test for AR(2)	1.79	p-value=0.074

As a contrast, we show the results without text features in Table 9. We can see that none of the intensity measures has a significant effect. This has two implications. First, without the text content, the mere counts of information do not affect sales directly, meaning consumers do read the content and make decision based on the textual information. Second, on average, the textual information from news articles or research articles has both favorable and unfavorable content to sales. Their effects balance out, so we do not observe a significant effect if we do not tease out the components (i.e., text features) for each piece of information. Furthermore, it is worth noting that the adjusted R-squared increases from 0.4182 to 0.7974 when we add the content features, indicating that the content features play a significant role in explaining consumer demand, above and beyond the information intensity. Therefore, the comparison between Table 8 and Table 9 shows the value of including text features in the analysis.

Figure 5. Box Plot of Credibility Scores from Survey

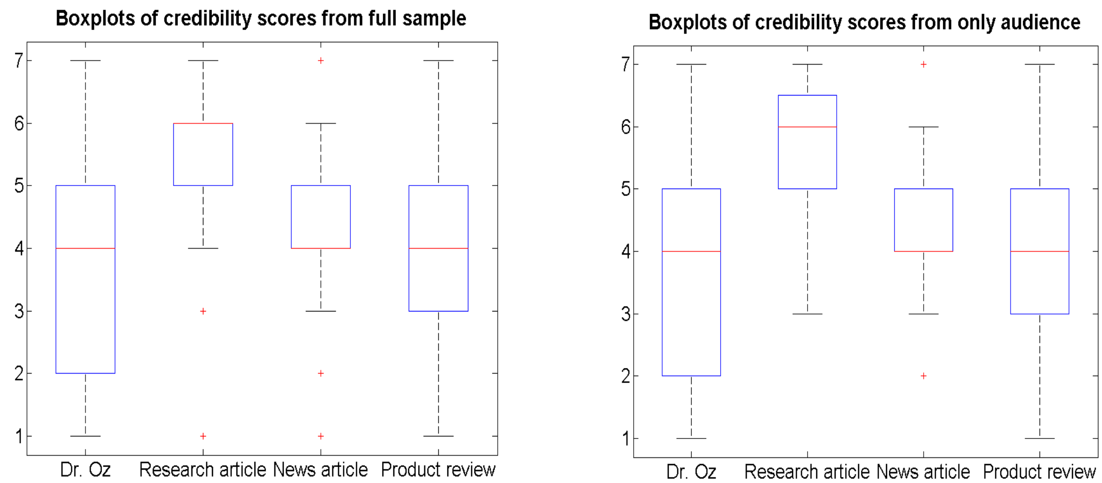


Table 10. Credibility Scores from Survey

	Mean	Std	Number of Responses
Dr. Oz Show	3.73	1.64	500
Research article	5.71	1.09	500
News article	4.39	1.12	500
Customer review	4.19	1.40	500

2.5.3 The Alternative "Oz Effect"

The evidence shown in Figure 1 indicates that, for products containing ingredients that were recommended on *The Dr. Oz Show*, there is an apparent growth in sales volume right after the show. However, our estimation results point to an explanation that involves no *direct* impact of *The Dr. Oz Show*. Given the widespread perception of Dr. Oz's big impact on healthcare products, we hypothesize that his show might have an indirect effect, which we label as the alternative "Oz Effect."

Our main results imply that news articles and scientific research articles have significant effects on sales through many language features. The data pattern in Figure 1 also shows an increasing number of news articles around the same time when the related weight-loss product went viral. This motivates us to explore the interaction among *The Dr. Oz Show*, news articles, research articles, and customer reviews. With the number of articles/shows/reviews from each source as a time series variable, we conduct a panel vector autoregressive (pVAR) analysis

(Abrigo & Love, 2015). On top of vector autoregressive model (VAR), pVAR estimates one single model using panel data. It is widely used to study the interdependencies of multiple time series. The pVAR model of order n, denoted by pVAR(n), is as follows:

$$\begin{aligned} news_{it} = {} & news_{it-1}\alpha_{11} + \cdots + news_{it-n}\alpha_{n1} + res_{it-1}\beta_{11} + \cdots + res_{it-n}\beta_{n1} + oz_{it-1}\gamma_{11} + \cdots \\ & + oz_{it-n}\gamma_{n1} + review_{it-1}\gamma_{11} + \cdots + review_{it-\mathrm{n}}\gamma_{n1} + u_{i,1} + e_{it,1} \end{aligned}$$

$$\begin{aligned} res_{it} = {} & news_{it-1}\alpha_{12} + \ldots + news_{it-n}\alpha_{n2} + res_{it-1}\beta_{12} + \ldots + res_{it-n}\beta_{n2} \\ & + oz_{it-1}\gamma_{12} + \ldots + oz_{it-n}\gamma_{n2} + review_{it-1}\gamma_{11} + \cdots + review_{it-n}\gamma_{n1} + u_{i,2} \\ & + e_{it,2} \end{aligned}$$

$$\begin{aligned} oz_{it} = {} & news_{it-1}\alpha_{13} + \ldots + news_{it-n}\alpha_{n3} + res_{it-1}\beta_{13} + \ldots + res_{it-n}\beta_{n3} \\ & + oz_{it-1}\gamma_{13} + \ldots + oz_{it-n}\gamma_{n3} + review_{it-1}\gamma_{11} + \cdots + review_{it-n}\gamma_{\square 1} + u_{i,3} \\ & + e_{it,3} \end{aligned}$$

$$\begin{aligned} review_{it} = {} & news_{it-1}\alpha_{13} + \ldots + news_{it-n}\alpha_{n3} + res_{it-1}\beta_{13} + \ldots + res_{it-n}\beta_{n3} \\ & + oz_{it-1}\gamma_{13} + \ldots + oz_{it-n}\gamma_{n3} + review_{it-1}\gamma_{11} + \cdots + review_{it-n}\gamma_{n1} + u_{i,4} \\ & + e_{it,4} \end{aligned}$$

where $i \in \{1,2,\ldots,N\}, t \in \{1,2,\ldots,T_i\}$, N is the total number of ingredients, u_i is the ingredient-level fixed effect, $news_{it}$ is the number of news articles on ingredient i and $news_{it-n}$ is the n^{th} lag value, similarly for res_{it}, oz_{it} and $review_{it}$. The $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}$ are the parameters to be estimated. Each observation is at ingredient-month level, because each article/show mentions ingredients rather than a specific product. The same ingredient applies to different product listings on Amazon, so the information intensities for products containing the same ingredient in each period are the same.

Table 11 displays the estimation results for pVAR(3).[31] There are in total 18 ingredients that all information sources covered during some overlapping periods, so we have 18 panels in total. $L1, L2, L3$ refer to the lag 1, lag 2, lag 3 values for each information source, accordingly.

The results tell us the number of The Dr. Oz Show in the last period is indeed positively correlated with the number of news articles in the current period. Though The Dr. Oz Show does not affect consumers' purchasing decisions directly, the evidence here supports its indirect impact on sales through stimulating more news articles on the related ingredient. This result tells us that, as a representative of celebrity doctors, Dr. Oz initiates some "hype news" that later on creates high waves of news articles. And it is the media-hype (Vasterman 2005; Zuckerman 2003), made possible by those celebrity doctors, that becomes highly influential on consumers' healthcare decisions.

[31] We examined models with lag variables of 1, 2, 3, 4, 5 lags, and so on. The results indicate no significant effects from the lag variables with order higher than 3. That is, there are significant results for 1, 2, 3, but not for higher-order lags.

Table 11. Panel Vector Autoregressive Analysis.

Variables	DV(1): #news	DV(2):#research	DV(3): #Dr. Oz Show	DV(4): #review
L1.#news	**0.322*** **	-0.0353	-0.00951	-179.0
	(0.0709)	(0.0706)	(0.0243)	(440.5)
L2.#news	-0.238	-0.00250	-0.00397	-82.44
	(0.739)	(0.0370)	(0.0109)	(196.7)
L3.#news	-0.271	-0.00969	-0.00385	-75.76
	(0.714)	(0.0292)	(0.0106)	(187.6)
L1.#research	-1.147	-0.143	-0.0126	-235.7
	(2.132)	(0.132)	(0.0318)	(556.7)
L2.#research	-1.425	-0.139	-0.0156	-357.5
	(3.373)	(0.165)	(0.0487)	(872.8)
L3.#research	**0.1847** **	0.0198	0.00229	-332.6
	(0.0850)	(0.152)	(0.0458)	(795.6)
L1.#Oz	**0.927** **	-0.355	-0.0949	-2,127
	(0.452)	(0.828)	(0.288)	(5,282)
L2.#Oz	-7.621	-0.155	-0.0927	-1,896
	(18.89)	(0.782)	(0.265)	(4,862)
L3.#Oz	-3.437	-0.0640	0.0638	-736.1
	(7.881)	(0.342)	(0.140)	(2,013)
L1.#Review	-0.0381	-0.0009	-0.0004	-8.080
	(0.0861)	(0.00350)	(0.00125)	(22.62)
L2.#Review	0.0138	0.000426	4.18e-05	3.134
	(0.0323)	(0.00124)	(0.000485)	(8.617)
L3.#Review	0.0123	0.000164	0.000220	2.965
	(0.0299)	(0.00123)	(0.000434)	(7.796)
Initial weight matrix	Identity			
GMM weight matrix	Robust			
No. of obs.	434			
No. of panels	18			
Avg. #periods[32]	26.1670			

Note: "DV" means the dependent variable in each pVAR equation; Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

This indirect Oz Effect has important policy implications. In addition to keeping an eye on what celebrity doctors say in public, policymakers should pay more attention to how healthcare-related information is relayed and propagated to the general public by news media. In other words, news articles (e.g., newspapers, magazines, etc.) are probably the true creators of the Oz Effect, amplifying, modifying, and distributing the show's message to a much broader audience. Moreover, we can see that research articles published three periods ago positively correlate with the current period's news intensity. It could be the case that journalist or news

[32] There are several reasons why the number of periods differs for different ingredients: the time when they entered the market is different; the time when they exited the market is different; some may have no sales during a period of time.

reporters, especially those specializing in healthcare market, would keep track of related research findings. When they read about some important research findings or see market interest about such findings, they will start writing news articles about them. In spite of the lag of three periods between research findings and news coverage, our main results in table 8 show that the research findings have an immediate impact on consumers' purchasing decision. These results indicate that many consumers stay tuned for the scientific findings on weight loss products, in a timelier manner than news reports.

In 2014, the Federal Trade Commission filed a lawsuit against a product that Dr. Oz peddled on his show. Senators criticized Dr. Oz during the congressional hearing: "I don't get why you need to say this stuff when you know it's not true. When you have this amazing megaphone, why would you cheapen your show? ... With power comes a great deal of responsibility." However, our results tell an alternative fact—that is, the sales boost may not be caused by Dr. Oz, but by public information from other sources. On the one hand, this relieves our concerns that some impactful celebrity doctors may mislead the public in order to achieve their own benefit. In fact, many news articles are taking roles of correcting or pointing out the misleading words from Dr. Oz[33]. On the other hand, to protect consumers, public policymakers should switch focus from a few individuals to other causes, in order to find out what information indeed plays an important role in consumers' purchasing decisions.

2.6 Conclusion

This paper aims to understand how public healthcare information affects consumers' healthcare choices. Multi-dimensional language features are extracted from textual information from different sources. Our textual analysis employs both traditional machine learning and state-of-the-art deep learning natural language processing techniques. We analyze thirteen textual features of large-scale textual information to understand whether these features play a significant role in consumers' healthcare choices. Among all the textual features, we include not only informative but also persuasive content features.

We contribute to the healthcare marketing literature by hugely improving the richness of information extracted from textual data and conducting structured analysis on consumers' healthcare choices. Our results have important managerial and public policy implications. Managerially, our findings provide guidance for managers on how to make effective advertising effort by incorporating favorable language features through different information sources. For public policymakers, our study sheds light on how to supervise health-related media content, in

[33] For example, McCoy, T. (2014). Half of Dr. Oz's Medical Advice is Baseless or Wrong, Study Says. Retrieved from https://www.washingtonpost.com/news/morning-mix/wp/2014/12/19/half-of-dr-ozs-medical-advice-is-baseless-or-wrong-study-says/?utm_term=.778f17a3c68d

order to protect consumers from misleading healthcare information and help their decision-making.

Surprisingly, our findings imply that *The Dr. Oz Show*, which is believed to be highly influential and potentially misleading in consumers' choice of healthcare products, does not significantly affect the sales of recommended products. News articles, as a fairly common media channel, are found to significantly affect the general public's choices in the healthcare domain. We also find that the availability and language features of the most credible information—scientific research articles—play an important role in affecting consumers' choices. As a result, we propose an alternative explanation of the "Oz effect." That is, instead of affecting sales directly, the show has an indirect impact on consumers' purchasing decision through stimulating more related news articles. *The Dr. Oz Show* is a typical example of TV shows launched or participated in by doctors. Therefore, our finding can demonstrate a general implication that healthcare information from this type of sources may initiate a media-hype, creating large waves of later-published news articles which tend to draw large sales volume to certain products.

Our analysis delivers several more interesting findings. We find that lowering a research article' complexity, ambiguity and emotion is especially effective in increasing the sales of a focal product. For product reviews, numerical ratings have a positive effect on sales but sentiment has a negative effect. Furthermore, we find that the product's sales increases when the language used in the product reviews is more ambiguous, subjective and positive. For both news articles and research articles, mentioning a product's name or performance has a significant positive effect on its sales.

There are a few limitations and directions for future research. First, this paper does not model the consumer information search process. With aggregate level data, our study focuses on the aggregate effect of each textual feature. We believe the aggregate-level findings are already rich, in terms of providing policymakers and managers valuable guidelines on their practices. However, future research might augment individual level data to explore consumer heterogeneity. Second, our research is most relevant to over-the-counter healthcare products. For prescription-required medicines, visiting physicians/doctors plays a big role. If access to physicians' prescription data is available, future research can be done on a broader range of healthcare products/services, for which other sources of information may have different effects.

Reference

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing*, 67(4), 103–117.

Bates, B. R., Romina, S., Ahmed, R., & Hopson, D. (2006). The effect of source credibility on consumers' perceptions of the quality of health information on the Internet. Medical Informatics and the Internet in Medicine, 31(1), 45–52.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research'. *Journal of Consumer Psychology*, 24 (4), 586–607.

Berger, J. and Milkman, K. L. (2012). What Makes online Content Viral? *Journal of Marketing Research*, 49 (2), 192–205.

Berger, J., Sorensen, A. T., Rasmussen, S. J. (2010). Positive Effects of Negative Publicity : When Negative Reviews Increase Sales. *Marketing Science*, *29*(5), 815–827.

Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., & Edu, J. B. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Chevalier, J. a, & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Ching, A. T., Clark, R., Horstmann, I., & Lim, H. (2015). The Effects of Publicity on Demand: The Case of Anti-Cholesterol Drugs. *Marketing Science*, (April), 150817104945009.

Chintagunta, P. K., Jiang, R., & Jin, G. Z. (2009). *Information, learning, and drug diffusion: The case of Cox-2 inhibitors*. *Quantitative Marketing and Economics* (Vol. 7).

D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, *125*(3), 26–33.

Dahlberg, L. (2001). Computer-Mediated communication and the public sphere: A critical analysis. *Journal of Computer-Mediated Communication*, *7*(1), 0.

Eisenberg, M. (2008). The Impact of Over the Counter (OTC) Weight Loss Product Television Advertisements on the Consumption of OTC Weight Loss Products and Other Diet-Related Behaviors. *Dissertation*.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal*, *324*(7337), 573.

Eysenbach, G. (2007). *Credibility of Health Information and Digital Media: New Perspectives and Implications for Youth*. *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the Helpfulness and Economic Impact of Product Reviews. *International Journal of Innovative Research and Development*, *23*(10), 1498–1512.

Goel, S., Watts, D., & Goldstein, D. (2012). The Structure of Online Diffusion Networks. *Proceedings of the 13th ACM Conference on Electronic Commerce*, 623-638.

Gong, J., Abhishek, V., & Li, B. (2017). Examining the Impact of Contextual Ambiguity on Search Advertising Keyword Performance : A Topic Model Approach. *MIS Quarterly*, (July), 1–68.

Hu, Y., & Sundar , S. S. (2010). Effects of Online Health Sources on Credibility and Behavioral Intentions. *Communication Research*, *37*(1), 105–132.

Huang, J.-H., & Chen, Y.-F. (2006). Herding in Online Product Choice. *Psychology & Marketing*, 23(5), 413-428.

Kalra, a., Li, S., & Zhang, W. (2011). Understanding Responses to Contradictory Information About Products. *Marketing Science*, *30*(December 2013), 1098–1114.

Khan, R., Misra, K., & Singh, V. (2016). Will a Fat Tax Work? *Marketing Science*, *35*(1), 10–26.

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*.

Lappas, T., Sabnis, G., & Valkanas, G. (2016). The Impact of Fake Reviews on Online Visibility : A Vulnerability Assessment of the Hotel Industry. *Information Systems Research*, *1*(December), 22.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Lee, D., Hosanagar, K., & Nair, H. S. (2017). Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook. Forthcoming, *Management Science*.

Lee, T. Y., & BradLow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, *48*(5), 881–894.

Liu, X., Lee, D., & Srinivasan, K. (2017). Large Scale Cross Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning. *Working Paper*, 1–48.

Metzger, M. J., & Flanagin, A. J. (2011). Using Web 2.0 Technologies to Enhance Evidence - Based Medical Information. *Journal of Health Communication: International Perspectives*, *16*(1), 45–58.

Porter, L. and Golan, G. J. (2006). From Subservient Chickens to Brawny Men: a Comparison of Viral Advertising to Television Advertising. *Journal of Interactive Advertising,* 6 (2).

Resnik, A., & Stern, B. L. (1977). An Analysis of Information Content in Television Advertising. *Source Journal of Marketing*, *41*(1), 50–53.

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*, 159–169.

Teisl, M. F., Bockstael, N. E., & Levy, A. (2001). Measuring the Welfare Effects of Nutrition Information. *American Journal of Agricultural Economics*, *83*(1), 133–149.

Timoshenko, A., & Hauser, J. R. (2017). Identifying Customer Needs from User-Generated Content. *Working Paper*.

Tirunillai, S. & Tellis , G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science,* 31(2), 198–215.

Uetake, K., & Yang, N. (2017). Success Breeds Success : Weight Loss Dynamics in the Presence of Short-Term and Long-Term Goals. *Working Paper*.

Vasterman, P. L. M. (2005). Media-Hype. *European Journal of Communication*, 20(4), 508–530.

Wang, P., Xu, J., Xu, B., Liu, C., Zhang, H., Wang, F., & Hao, H. (2015). Semantic Clustering and Convolutional Neural Network for Short Text Categorization. *Proceedings ACL 2015*, 352–357.

Ye, Q., Law, R., & Gu, B. (2009). International Journal of Hospitality Management The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*, 180–182.

Yildirim, P., Gal-Or, E., & Geylani, T. (2013). User-Generated Content and Bias in News Media. *Management Science*, *1909*(January 2015), 1–12.

Zhang, Y., & Wallace, B. (2016). A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. *arXiv Preprint*, arXiv:1510.03820v4.

Zuckerman, D. (2003). Hype in health reporting: "checkbook science" buys distortion of medical news. *International Journal of Health Services*, *33*(2), 383–389.

Appendix

2.7 Survey Form Designed in Amazon Mechanical Turk

We are studying textual information about weight-loss products.

You will see a text. The following questions ask if a certain type of content is present in the review.

Please read the text and questions carefully before answering the questions.

QUESTIONS:

Please answer the following questions about the text. Each one has a yes/no question and some have follow-up questions.

Does this text:

1. *[Side Effect]* contain content regarding **side effect** of the product?

(e.g., "It doesn't have any side effect on me.", "It causes some pretty painful diarrhea.")

Yes ◯ No ◯

If you answered yes above, judge if the product has a side effect. If answered no above, then select Not Applicable.

It has side effect ◯ It has NO side effect ◯ Not Applicable ◯

2. *[Performance]* contain content regarding the product's **performance**?

(e.g., helped losing weight/suppress appetite/increase metabolism, etc.)

Yes ◯ No ◯

If you answered yes above, judge if the product is effective in helping losing weight. If answered no above, then select Not Applicable.

It is effective ◯ It is NOT effective ◯ Not Applicable ◯

3. *[Target]* is targeted towards some specific consumer segment(s)? E.g., particular gender/age/location or people with certain qualifications or characteristics.

(e.g., "Strawberry lovers, go get this one.")

Yes ◯ No ◯

4. *[Price]* Does this text mention price or make a price comparison against other products?

Yes ◯ No ◯

If you answered yes above and there is a price comparison, judge if it is a favorable comparison(i.e., the product's price is lower than that of others.).

If answered no above or there's no price comparison, then select Not Applicable.

A favourable comparison ◯ Not a favorable comparison ◯ Not Applicable ◯

5. *[Deal]* Does this text contain information about any type of promotions? E.g., discounts, deals, coupons, free items, etc.

Yes ◯ No ◯

6. *[Product Name]* Does this text mentions a product name or brand name?

Yes ◯ No ◯

7. *[Humor]* Does this text use humor?

Yes ◯	No ◯

8. *[Emotion]* Does this text include emotional content? (e.g., "I feel so sad that it has no effect on me.")

Yes ◯	No ◯

If you answered yes above, judge if it is a positive or negative emotion. If answered no above, then select Not Applicable.

Positive emotion ◯	Negative emotion ◯	Not Applicable ◯

2.8 B. Journal Names

Source of Research Articles: Peer Reviewed Journals

JOURNAL NAME	JOURNAL NAME	JOURNAL NAME	JOURNAL NAME
Journal of Materials Science : Materials in Medicine	Emirates Journal of Food and Agriculture	Clean Technologies and Environmental Policy	Australian Journal of Herbal Medicine
PLoS One	The Journal of Dairy Research	Journal of Biological Engineering	Biodegradation
International Journal of Molecular Sciences	Critical Reviews in Food Science and Nutrition	European Food Research and Technolog	Anti - Corrosion Methods and Materials
Marine Drugs	Gene Therapy	Virology Journal	Molecular Biology Reports
Journal of Nanomaterials	Indian Journal of Pharmaceutical Sciences	Polymer Composites	Antonie van Leeuwenhoek
Journal of Food Science and Technology	Tropical Animal Health and Production	Journal of Bone and Mineral Metabolism	Oxidative Medicine and Cellular Longevity
Lipids	Carcinogenesis	Nutrition Research Reviews	PPAR Research
International Journal of Polymer Science	Journal of Industrial Microbiology & Biotechnology	Current Topics in Nutraceuticals Researc	AIDS Research and Therapy
Journal of Polymers and the Environment	Obesity	PLoS Neglected Tropical Diseases	Pakistan Journal of Zoology
Pharmaceutical Research	The Proceedings of the Nutrition Society	Public Health Nutrition	Al Ameen Journal of Medical Sciences
The British Journal of Nutrition	International Journal of Environmental Science and Technology	International Journal of Obesity	Biomaterials Research
Journal of Dairy Science	International Journal of Food Science and Technology	Polymer Engineering and Science	Notulae Scientia Biologicae
Journal of Polymer Materials	Journal of Food Protection	Kidney International	Biomedical Microdevices
Scientific Reports (Nature Publisher Group)	Journal of Food Science	Diabetology & Metabolic Syndrome	Advances in Materials Science and Engineering
BioMed Research International	Journal of Dentistry	Diabetes	Biophysics
African Journal of Biotechnology	Tissue Engineering	Water Environment Research	Acta Agron—mica
Vaccine	International Journal of Biomaterials	Agroforestry Systems	Acta Chimica Slovaca
Nanomedicine	Molecular Therapy	Journal of Nanobiotechnology	BioResearch Open Access

Source of Research Articles: Peer Reviewed Journals (Continued)

JOURNAL NAME	JOURNAL NAME	JOURNAL NAME	JOURNAL NAME
Nanomedicine	Molecular Therapy	Journal of Nanobiotechnology	BioResearch Open Access
Environmental Science and Pollution Research International	Chemistry Central Journal	Analytical and Bioanalytical Chemistry	Journal of the International Society of Sports Nutrition
Journal of Animal Science	Journal of Young Pharmacists	Clinical Lipidology	BMC Plant Biology
Applied Mechanics and Materials	World Journal of Life Sciences and Medical Research	Pigment & Resin Technology	BMC Genomics
International Journal of Pharmaceutical Sciences and Research	Environmental Chemistry Letters	Journal of Nanotechnology	The Journal of Nutrition
Bioprocess and Biosystems Engineering	Biotechnology Letters	Alternative Therapies in Health and Medicine	Journal of Strength and Conditioning Research
Nutrition	International Journal of Plastics Technology	Microbial Cell Factories	Ethiopian Journal of Environmental Studies and Management
Obesity Research	Caries Research	PLoS Pathogens	Journal of Nutrition and Metabolism
BMC Complementary and Alternative Medicine	Cancer Letters	Pakistan Journal of Medical Sciences Quarterly	Gastroenterology Research and Practice
International Food Research Journal	Acta Pharmaceutica	Research in Veterinary Science	Gesunde Pflanzen
Applied Microbiology and Biotechnology	Annals of Biomedical Engineering	Nutrition Reviews	Journal of Applied Poultry Research
European Food Research and Technology	Archives of Virology	Nutrition Journal	Euphytica
Applied Biochemistry and Biotechnology	European Journal of Plant Pathology	Nutrition and Cancer	Archives of Pharmacy Practice
European Journal of Nutrition	Journal of Biomedicine and Biotechnology	Nutrition & Metabolism	Amino Acids
Applied Biochemistry and Microbiology	Materials Science and Technology	Research Journal of Pharmacy and Technology	Annals of Nutrition & Metabolism
Indian Journal of Clinical Biochemistry	Journal of Conservative Dentistry	Nephron	International Journal of Cultural Property
Biotechnology and Bioprocess Engineering : BBE	International Journal of Carbohydrate Chemistry	Mycorrhiza	Agricultural History

Source of Research Articles: Peer Reviewed Journals (Continued)

JOURNAL NAME	JOURNAL NAME	JOURNAL NAME	JOURNAL NAME
In Vitro Cellular & Developmental Biology	Journal of Membrane Biology	Polish Journal of Veterinary Sciences	Journal of Insect Behavior
Global Journal of Research on Medicinal Plants	Journal of International Dental and Medical Research	Journal of Applied Microbiology	Age
Lipids in Health and Disease	Asian Journal of Research in Chemistry	Metabolic Brain Disease	International Journal of Food Sciences and Nutrition
Molecular Medicine Reports	Bulletin of Experimental Biology and Medicine	Journal of Sports Medicine and Physical Fitness	International Journal of Molecular Medicine
Iranian Journal of Basic Medical Sciences	Biological Trace Element Research	The Veterinary Record	Maejo International Journal of Science and Technology
The Pharma Innovation	Journal of Electronic Materials	Journal of Mammary Gland Biology and Neoplasia	Human and Experimental Toxicology
International Journal of Obesity and Related Disorders	Evidence-Based Complementary and Alternative Medicine	Science International	Journal of Toxicology
Asian Journal of Pharmaceutics	Experimental and Therapeutic Medicine	Journal of Diabetes & Metabolic Disorders	Italian Journal of Food Science
European Journal of Clinical Nutrition	Journal of Basic and Clinical Physiology and Pharmacology	Journal of Clinical Pharmacy and Therapeutics	Journal of Pharmaceutical Sciences and Research
Animal: an International Journal of Animal Bioscience	Journal of Pharmaceutical Education and Research	Journal of Bioenergetics and Biomembranes	Journal of Ayurveda and Integrative Medicine
Plant Growth Regulation	Journal of Materials Research	Focus On Geography	Journal of Biological Research
The American Journal of Surgery	International Journal of Photoenergy	Journal of Medical Toxicology	Theatre Research International
The Scientific World Journal	Metabolomics	Poultry Science	The Protein Journal
Food Biophysics	Circulation	BMJ Open	Journal of Mountain Science
Polymers & Polymer Composites	Pharmacognosy Communications	Asia Pacific Journal of Clinical Nutrition	Annual Review of Nutrition
Malaysian Journal of Pharmaceutical Sciences	Journal of Ocean University of China. JOUC	American Journal of Rhinology & Allergy	Journal of the American Dietetic Association
International Journal of Applied Science and Engineering	Fish Physiology and Biochemistry	Current Topics in Nutraceuticals Research	Breast Cancer Research and Treatment
Diabetes Care	Glycoconjugate Journal	Defence Science Journal	Diabetologia
Indian Journal of Medical Research	The Journal of Agricultural Science	International Journal of Aquaculture	International Archives of Allergy and Immunology

2.9 C. Results for model with lags of text features

VARIABLES	Estimates	VARIABLES	Estimates	VARIABLES	Estimates
L.sale	0.620***	smog_news	-0.230	performp_res	0.108
	(0.0797)		(0.253)		(1.739)
price	-0.00174	L.smog_news	-0.714***	L.performp_res	-3.551**
	(0.0204)		(0.261)		(1.724)
avg_rate	-2.136	smog_rew	-0.0877	performp_news	-0.864
	(1.787)		(0.183)		(1.474)
n_ads	-0.783	L.smog_rew	0.129	L.performp_news	0.778
	(2.490)		(0.0855)		(1.432)
intensity_oz	-16.94	subj_oz	-18.04	performp_rew	3.272**
	(33.06)		(11.83)		(1.623)
L.intensity_oz	40.39*	L.subj_oz	3.371	L.performp_rew	0.0521
	(24.22)		(7.601)		(0.450)
intensity_res	0.904	subj_res	1.865	performn_res	5.181*
	(1.538)		(2.497)		(2.904)
L.intensity_res	-2.050	L.subj_res	-3.456	L.performn_res	-2.218
	(2.493)		(2.276)		(2.079)
intensity_news	0.00328	subj_news	1.963*	performn_news	-20.76
	(0.117)		(1.178)		(15.34)
L.intensity_news	0.232*	L.subj_news	-2.209	L.performn_news	-9.365
	(0.129)		(1.780)		(9.826)
ambiguity_oz	-9.489	subj_rew	-0.849	performn_rew	-0.821
	(7.252)		(0.768)		(1.041)
L.ambiguity_oz	18.94	L.subj_rew	-0.568*	L.performn_rew	0.0317
	(18.85)		(0.320)		(0.590)
ambiguity_res	-1.698	price_news	-1.982	pname_oz	39.44
	(1.829)		(1.658)		(25.73)
L.ambiguity_res	-0.517	L.price_news	3.545**	L.pname_oz	6.346
	(2.388)		(1.729)		(11.91)
ambiguity_news	4.497***	price_rew	0.982	pname_res	3.281
	(1.520)		(1.689)		(2.288)
L.ambiguity_news	1.282	L.price_rew	-0.262	L.pname_res	-2.410
	(1.500)		(0.904)		(1.818)
ambiguity_rew	0.268	emotionp_res	-4.170*	pname_news	10.81***
	(0.810)		(2.350)		(3.313)
L.ambiguity_rew	0.140	L.emotionp_res	-1.684	L.pname_news	6.053**
	(0.479)		(1.827)		(2.888)
sent_res	25.18*	emotionp_news	-0.448	pname_rew	-0.641
	(15.28)		(1.055)		(1.435)
L.sent_res	13.27	L.emotionp_news	-0.0427	L.pname_rew	-0.796
	(10.44)		(1.264)		(0.783)
sent_news	17.10*	emotionp_rew	0.874	hum_rew	-18.00
	(8.810)		(1.034)		(25.83)
L.sent_news	1.533	L.emotionp_rew	-0.665	L.hum_rew	9.301
	(7.680)		(0.601)		(7.434)
sent_rew	-3.462*	emotionn_news	1.037	side_news	4.924
	(1.943)		(1.937)		(7.381)
L.sent_rew	-0.434	L.emotionn_news	4.302*	L.side_news	15.97
	(0.774)		(2.240)		(13.03)
smog_oz	-7.811**	emotionn_rew	-1.392	side_rew	-5.255*
	(3.465)		(1.221)		(2.704)
L.smog_oz	-4.584*	L.emotionn_rew	0.792	L.side_rew	-0.495
	(2.580)		(0.654)		(1.048)
smog_res	-0.876**	L.smog_res	0.0941		
	(0.366)		(0.403)		
Observations			4,965		
R-squared			0.852		

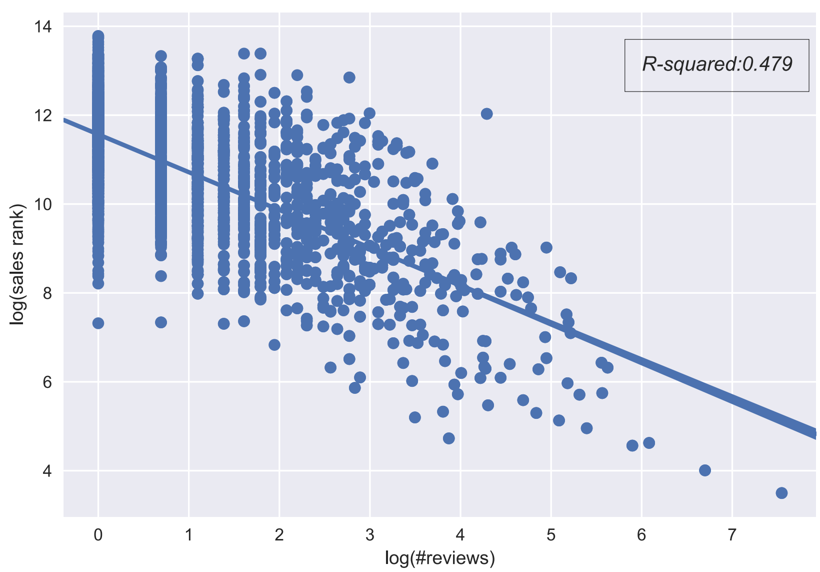
Robust standard errors in parentheses

**** p<0.01, ** p<0.05, * p<0.1*

2.10 E. Correlation between ln(sales rank) and number of reviews

Mayzlin & Chevalier (2006) established a linear relationship between ln(sales) and ln(sales rank), supporting using sales rank as a valid proxy measure for sales. To further check the validity of our proxy measure (i.e., number of reviews), we check the relationship between ln(sales rank) and ln(number of reviews) with the cross-sectional data of sales ranking within-category (i.e., healthcare category) on Amazon.com. Figure E1 shows the linear fit between ln(sales rank) and ln(number of reviews), using the cross-sectional data of sales ranking. We can see that in our data ln(sales rank) and ln(number of reviews) also have a linear relationship, and they are highly correlated with a negative correlation -0.692. This further supports the use of review count as a proxy measure of sales.

Figure E1. Linear Fit



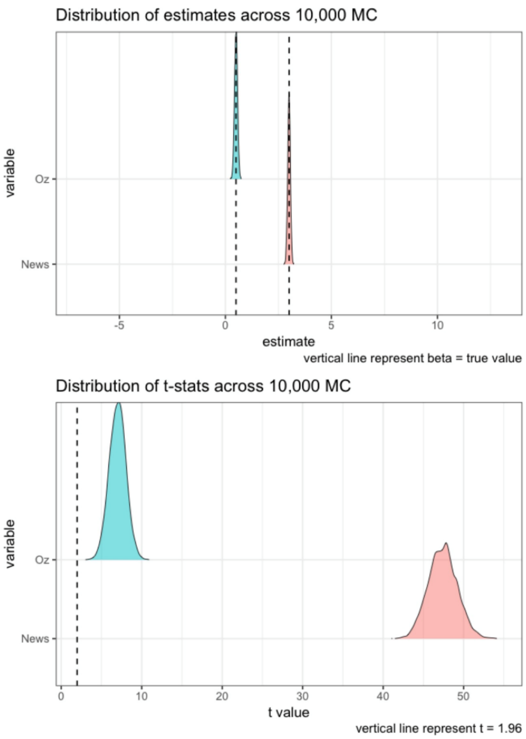
2.11 F. How measurement errors in sales measure affect the results?

Our main findings indicate that *The Dr. Oz Show*, which is believed to be highly influential and potentially misleading in consumers' choice of healthcare products, does not significantly affect the sales of recommended products. News articles and research articles are found to significantly affect the general public's choices in the healthcare domain. A following up question is, whether the insignificance effect of hype news (Dr. Oz Show) is driven by the measurement error in our sales measure (i.e., number of reviews in each period). Let us use a simplified model to investigate this question.

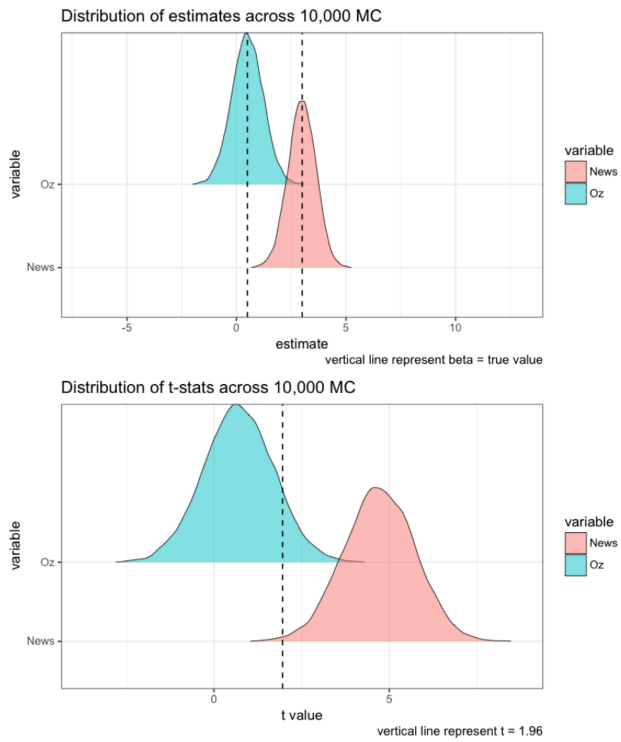
Suppose we have the model $Y = \alpha_0 + \alpha_1 Intensity^{Oz} + \alpha_2 Intensity^{News} + \text{e}$. For variables $Intensity^{Oz}$ and $Intensity^{News}$, we can draw them with means 1.11 and 68.28 to keep consistent with the summary statistics in table 4, and assume each has a variance of 1. Because a larger measurement error in Y equivalently means a larger variance of e, we consider two cases: Var(e)=20, and Var(e)=2. With some pre-set values of $\boldsymbol{\alpha}$**,** we can run 10,000 MC for this data generating process with 1000 data points. Then we can estimate the model with the resulting data and get the distribution of parameter estimates for $\boldsymbol{\alpha}$, as well as the t-statistics. We are interested to see under what circumstances the estimate $\widehat{\alpha}_1$ is significant when Var(e)=2 but becomes insignificant under Var(e)=20, whereas the estimate $\widehat{\alpha}_2$ remains significant in both cases. We found that this could happen only when the true parameter $\alpha_2 > \alpha_1$. In Table F1, we show the results for the two cases with true parameters $\alpha_0 = \alpha_2 = 3$, and $\alpha_2 = 0.5$. The results indicate that, for all simulations both Oz and News effects are significant when Var(e)=2. However, when Var(e)=20, for 99.7% simulations the News effect is significant but for 85.5% of simulations the Oz effect is insignificant at a 95% confidence interval.

In other words, measurement error may cause a significant Oz effect to become insignificant only if the news articles have a bigger impact on sales than the Dr. Oz Show does. This is exactly consistent with our main finding: Dr. Oz is not that influential as many people thought he was; in fact, news articles play a more important role in steering consumers' healthcare choices.

Table F1. Estimation Results with Simulated Data



(a) Simulation results when Var(e)=2



(b). Simulation results when Var(e)=20

Chapter 3
Does Fast Fashion Increase the Demand for Premium Brands? A Structural Analysis

3.1 Introduction

The global fashion industry has reached an estimated value of 3 trillion dollars.[34] Traditional luxury fashion brands, such as Gucci, Prada, and Louis Vuitton, have maintained a strong position within the industry, backstopped by the increasing demand from developing economies such as China. At the same time, fast fashion brands such as Zara, Forever 21, and H&M have been storming the globe with their versatile styles and low price. The path to success of these fast fashion brands, however, is nothing short of controversy. Every year, large fast fashion chains spew close-to-the-runway originals at lightning speed. On the one hand, the high-end brands, believing these copycats will steal their customers and hurt their profitability, spare no effort in fighting back by launching lawsuits against them.[35] On the other hand, high-end brands may not face any threat if their consumers and those of fast fashion brands are different segments with variant values of brands and styles. In addition, lawmakers tend to view the utilitarian nature of clothing and fashion as more important than its artistic and stylistic purposes; therefore fashion designs are not under the protection of copyright law.[36] In spite of this tension on the enforcement and effects of copycats, the effect of fashion copycats on high-end brands remains empirically unclear.

In this paper, we estimate the impact of low-end copycats on the demand for high-end brands. We develop a dynamic structural model of individual consumer's fashion choices, which allows for counterfactual analysis of alternative copyright policies against copycats. Contrary to the conventional wisdom, we find that prohibiting low-end copycats can decrease the demand of high-end brands significantly. We find novel mechanisms contributing to this result, which are distinct from the promotional effect documented in the counterfeit literature (e.g., Qian, 2014): first, fewer style choices from low-end brands would limit the mix-and-match choices for consumers and put them on greater financial constraint to get a satisfactory ensemble of clothes, resulting in them buying less high-end brands; second, the lack of good styles from low-end brands will make it harder for consumers to build up their popularity/likeability, which limits the complementary value of high-end brands. As a result, consumers adopt less high-end brands. Our findings suggest that the above-mentioned market expansion effect dominates the

[34] https://fashionunited.com/global-fashion-industry-statistics

[35] For example, https://wwd.com/business-news/legal/the-5-five-biggest-lawsuits-facing-fashion-retail-10875211/

[36] http://www.thefashionlaw.com/home/how-do-fast-fashion-retailers-get-away-copying-high-fashion-brands

competition effect. Other counterfactual analyses examine the consequence if the brand or peer feedback cannot be seen, which is the case for many other social medium and offline markets.

We overcome substantial technical and empirical challenges to obtain these results. Traditionally, there are two challenges to studying the micro-level consumer choices of fashion goods. First, fashion styles are not quantifiable. Second, individual-level choices on fashion brands and styles across a large pool of brands are not available. In this paper, we employ state-of-the-art deep learning techniques to quantify fashion styles from fashion images. We overcome the second challenge by studying the choices over brands and styles for fashion conscious consumers on social media. Nowadays, fashion is one of the most popular contents generated by users on social media (Hu et al. 2014). More and more fashion consumers post what they wear online, and importantly, these social media users become trendsetters and influence a large number of other fashion consumers.[37] Therefore, investigating how these consumers make choices on brand and style can help us understand the market demand of fashion goods. Our data is from a large online fashion-sharing platform where users post their fashion pictures and evaluate others' pictures. The data comprise 10262 active users and 64681 fashion posts and span over three years. We account for consumer heterogeneity and estimate the structural model following a Hierarchical Bayesian framework.

Substantively, our results have implementable policy implications to both managers as well as policymakers. Managerially, we provide novel insights on how copycats can help the high-end brands, which guide their product strategy. In fact, some high-end brands have started to produce their own low-end frugal version of similar styles, consistent with the first mechanism of copycat effects in our findings. Moreover, for fashion companies, understanding how fashion consumers value brands and styles can help managers infer the market demand and make the optimal investments in branding and product design. From the policy-making perspective, we provide novel insights on the potential consequence of alternative copyright policies for fashion designs. More generally, our findings speak to the debate on whether copyright or patent protection encourages or discourages innovation in the fashion industry: with more demand brought by low-end copycats, companies can get more money to invest in innovation, which may lead to more creative designs for the entire fashion market.

This paper also contributes methodologically and theoretically. Methodologically, we make two contributions. First, we develop a framework to analyze consumer choices where visual features are important product attributes and other people's opinions heavily affect the decision-making. Second, we use deep learning and image processing techniques to quantify fashion styles to make the analysis of fashion style choices possible. Theoretically, our findings

[37] http://www.latimes.com/fashion/la-ig-bloggers-20160809-snap-story.html

provide new insights on how copycat products can benefit the original brands, which also apply to the cases of counterfeits and pirated goods if consumers have mix-and-match choices and popularity concern.

The rest of the paper is organized as follows. In Section 2, we review the literature related to this paper. Section 3 presents the raw data, visual feature extraction, and exploratory analysis. Section 4 describes our model. We illustrate the identification and estimation strategy in Section 5. We report the estimated results in Section 6, followed by the counterfactual analysis in Section 7. Section 8 concludes.

3.2 Literature Review

Our study relates to marketing and economics literature on branding, counterfeits and piracy, conspicuous good consumption, as well as the literature on machine learning methods and applications.

As we seek to examine the brand value and style value for fashion goods, our paper is related to the marketing literature on branding (e.g., Borkovsky et al., 2017; Goldfard, 2009; Keller & Lehmann, 2006; Kamakura & Russell, 1993). More recently with unstructured data, Nam et al. (2017) investigate the qualitative brand information harvested from social tags in the textual form. Liu et al. (2018) study consumers' brand perception on social media using visual data. In this paper, we focus on fashion goods, specifically clothing. We examine how consumers value brands versus styles, and how copycat styles affect the demand of premium brands measured by the units of clothing items adopted in the social media posts.

This paper relates to the literature of counterfeits and piracy (e.g., Qian, 2014; Ma et al., 2016; Oberholzer-Gee & Strumpf, 2007; Smith & Telang, 2009), which provided evidences of both cannibalization and promotional effects of counterfeits (pirated goods) on the original. In contrast, copycats and counterfeits are fundamentally different. Counterfeits copy not only the style or content but also the trademark (i.e., the brand logo), therefore they violate the trademark law. However, copycats do not copy the brand logo and are typically legal. Therefore, counterfeits can benefit the original brand by improving the awareness of the brand (i.e., promotional effect), but copycats cannot directly give consumers information about the original brands. Studies that examine the market response of copycats include Horen and Pieters (2012) and Wang et al. (2018). Horen and Pieters (2012) conducted lab experiments and survey studies at a grocery context to demonstrate how copycats can gain or lose from their resemblance to the original brands, but they remain silent about how copycats affect the demand of the original brands. Wang et al. (2018) examine the aggregate impact of copycat mobile apps on the demand of original apps. They find that deceptive and low-quality copycat apps may positively affect the demand of the original app, implying the existence of the promotional effect. In contrast, fashion goods are fundamentally different from grocery and mobile apps, in the sense that consumers mix and match multiple clothing items and peer feedback plays an important role. Moreover, as will be shown in this paper, our micro-level study specifies new mechanisms on how copycats can benefit premium brands, unlike the traditional promotional effect.

Various theoretical works have investigated fashion firms' strategies on information disclosure (e.g., Yoganarasimhan, 2012), competitive pricing (e.g., Amaldoss & Jain, 2005), given consumers' dual needs of conformity and differentiation in conspicuous consumption (Brewer, 1991). Accordingly, the consumer tradeoff between expressing individuality and conforming to others' opinions/likes, related to self and public self-consciousness for social behaviors in psychology (Fenigstein et al., 1975). Though people have heterogeneous underlying preferences, Bernheim (1994) shows that when status is very important relative to intrinsic utility, some people will conform to a single standard of behavior. To our knowledge, this is the first empirical research examining consumers' conspicuous consumption while incorporating both their intrinsic preferences and the impact of others' opinions.

To extract and quantify the styles of fashion goods, we need to analyze visual data by referring to machine learning literature for image analytics. Specifically, we apply support vector machine (SVM), support vector regression (SVR), Fast R-CNN (Girshick, 2015), and transfer learning using Siamese CNN (Hadsell et al., 2006; Veit et al., 2015) to extract clothing style features (i.e., compatibility and distinctiveness) and user appearance features (i.e., facial attractiveness and body BMI).

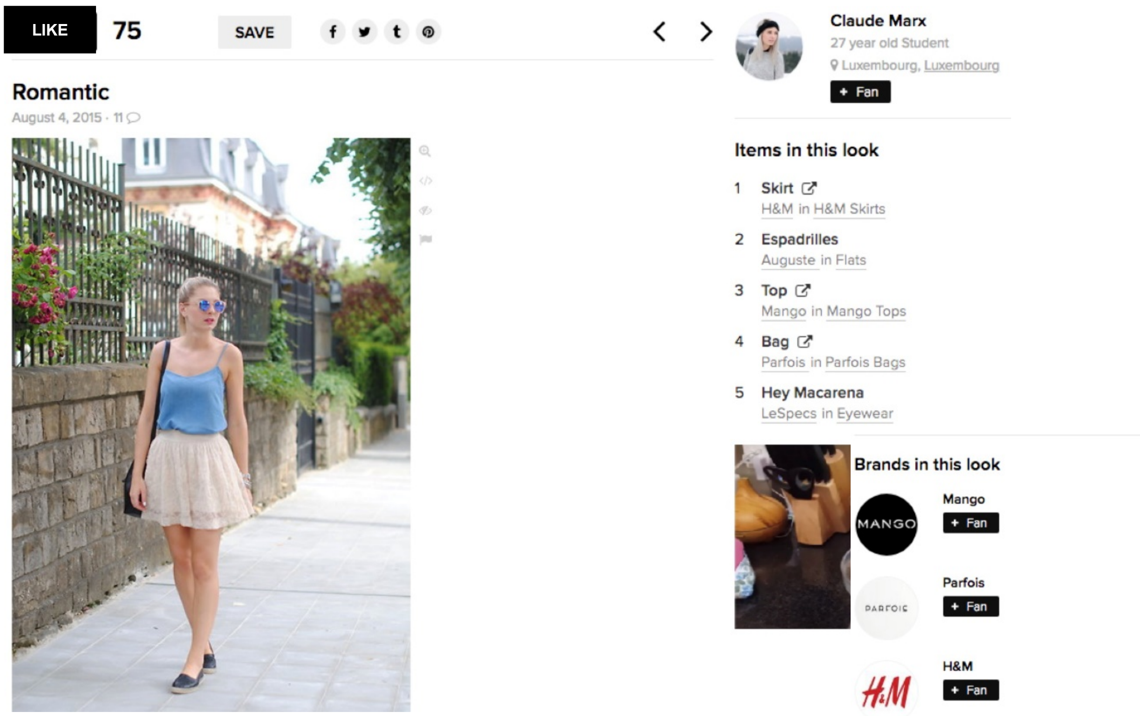
3.3 Data

The research context in this paper is the world's largest online fashion sharing community,[38] designed for users to post their own fashion photography, featuring themselves and their outfits. It shares similar features with other photo-based social media except that the content is restricted to fashion. More importantly, the website features a special function: a user tags the brand of the fashion items in his/her posted picture. Therefore, the brand information is clearly listed beside the fashion look[39] and can be seen by others. Figure 1 shows what a fashion post looks like on the website.

[38] The company was launched in 2008. As of July 2017, there are more than 6 million users registered.

[39] We use "fashion look" to refer to a picture.

Figure 1. Example of a Fashion Post



We collect individual-level historical data from August 2013 to August 2017. The data set contains the entire history of fashion content generation for a random sample of 10,262 users[40] who registered after August 2013 and posted at least once. For each fashion look, we collect the image data, the brands for the clothing items, the time stamp, as well as how many likes the picture has attracted. For each user, we also observe his/her age from the brief biography. The gender information is not directly observable, and we will predict the gender from their picture in section 3.1.2.

Table 1. Descriptive Statistics

(a) User-level summary statistics

	#posts	#Cumulative likes	#Following	Age
mean	11.3242	173.1065	31.0892	23.4031
Std.	16.8707	1476.7562	116.6762	4.7575
min	1	0	0	4
Median	5	25	14	24
max	561	93376	6130	99

[40] We use "fashion bloggers" and "users" interchangeably throughout the paper.

(b) Post-level summary statistics

	Mean	Std.	Min	Median	Max
#likes	25.8841	52.1091	0	10	1747

Note: "#" denotes "the count of."

Table 1 shows the summary statistics for these users and their fashion posts. We can see that the standard deviation is large relative to the means, and the measures have skewed distribution. A small group of people has lots of posts and likes, whereas many others post very few. This observation is similar to that of most social media platforms.

Brand Categorization. Following Ha et al. (2017), we group the fashion brands into three categories: fast fashion (high street), designer, and mega couture.[41] The categorization is according to domain experts in the fashion industry, based on brand identity and price ranges. We show some examples of each brand category in Table 2.

Table 2. Brand Categorization

Brand Categories	Examples
Level 1: Fast Fashion (High Street)	Zara, H&M, Forever21
Level 2: Designer	Kate Spade, Coach, Michael Kors
Level 3: Mega Couture	Gucci, Prada, Chanel

3.3.1 Feature Extraction from Images

For a fashion look, we focus on two key aspects of visual features that can affect one's utility: the clothing styles and the appearance of the users.[42] We describe how we extract and measure the clothing styles in 3.1.1 and user appearance in 3.1.2.

3.3.1.1 Clothing Styles

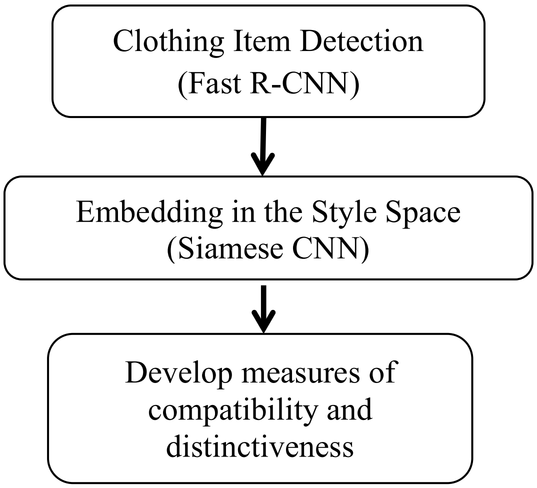
In light of fashion satisfying consumers' social needs for group cohesion and differentiation (Simmel, 1904) and domain experts' opinion from well established fashion magazines, at a high level, two style features that are particularly relevant for fashion goods: compatibility and distinctiveness. Compatibility speaks to the combination of clothing items from different categories (e.g., shirts versus pants), whereas distinctiveness measures how visually differentiated each item is from others within the same category. We abstract away from more

[41] The original categorization also includes "small couture" brands, but there is only one observation of such brand in our data. So we consider only three categories.

[42] On the blogging platform, the user himself/herself is the model in the picture. One account consistently posts the account owner's fashion look.

granular style factors (e.g., color, texture) and capture the styles at a high level, because those granular factors can also be described or evaluated according to compatibility and distinctiveness. Below we explain how we extract these two style features from the fashion looks.

Figure 2. Steps of Clothing Style Features Extraction



To measure fashion styles, we first need to detect or identify the clothing items in each fashion look. We follow the approach of the DeepFashion project by Liu et al. (2016). The method is based on the application of Fast Region-based Convolutional Neural Network (Fast R-CNN) (Girshick, 2015). A Fast R-CNN network takes an image and a set of object location proposals as inputs. It learns to classify objects and refine their spatial locations jointly. We adopt the network architecture of DeepFashion, which was trained on the largest and most comprehensive clothes dataset to date, annotated with clothing landmarks and categories.

For each fashion look, we extract only the clothing items, which are the most visually dominating items in a picture. Most accessories are too small to be precisely detected, so we do not include them in the analysis in this paper. We keep the cropped items (upper and bottom) if the confidence scores are higher than some threshold.[43] If the detector cannot separate the top and bottom items, we treat the clothes as full-body outfits whose compatibility is assigned an average score of the fashion looks posted during the past three months (same length as a season).

After we detect the clothing items, we can proceed to measure compatibility and distinctiveness for the fashion looks.

---

[43] We tried 0.7, 0.8, and 0.9 for robustness checks.

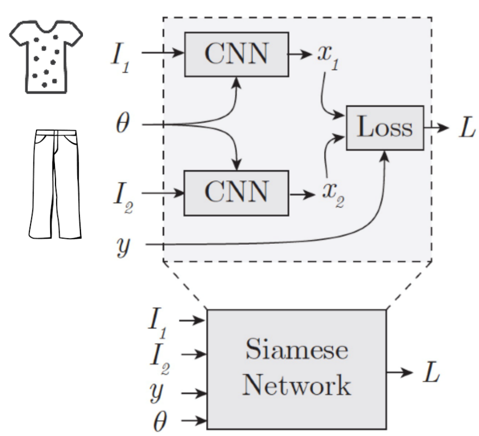
3.3.1.1.1 Compatibility

Among a large number of fashion looks, we would like to know what clothing items go well together. We adopt a deep learning approach to learn a feature transformation from images of clothing items into a latent space that represents compatibility. We use a Siamese convolutional neural network architecture (Siamese CNN) (Hadsell et al., 2006), where training data are pairs of items that are either incompatible or compatible.

To measure compatibility, we first initialize the model with weights trained on two million pairs of labeled pairs collected from the purchase data of fashion goods on Amazon.com (Veit et al., 2015). As purchasing two items together may not necessarily mean the consumers treat the items as compatible, we further collect an additional training dataset by conducting a survey on Amazon Mechanical Turk (henceforth AMT). We directly ask survey respondents' opinion on compatibility of randomly selected pairs of items. With transfer learning, we fine-tune the deep neural network with three thousand pairs of responses (compatible versus incompatible) from the survey to improve the measure of compatibility. Please see the appendix for the survey design.

The abstraction of Siamese CNN architecture is shown in Figure 3. Essentially, it learns a feature transformation $f: I \rightarrow X$ from the image space I (i.e., raw representation of images) to the style space X (i.e., another representation that captures the style features). In the style space, compatible items are closer together, and incompatible items are farther away. Then, we can use the distance between two items' locations in the style space to measure how compatible they are. In Figure 3, I_1 and I_2 are the inputs of two clothing items from different categories (top and bottom), x_1 and x_2 are vector representations in the style space, y is the label of data (either compatible or incompatible), and θ is the set of parameters that specify the neural network, which we need to estimate.

Figure 3. Abstraction of Siamese CNN Architecture



The loss function $L(\theta)$ is a contrastive loss and can be expressed as:

$$L(\theta) = \sum_{(x_1, x_2)} L_p(x_1, x_2) + \sum_{(x_1, x_3)} L_n(x_1, x_3)$$

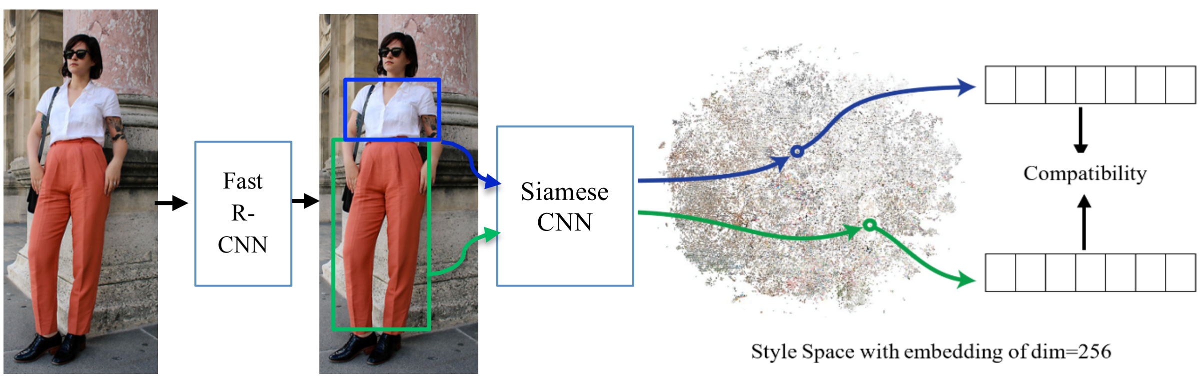
The first term L_p penalizes when a compatible pair is too far apart, and the second term L_n penalizes when an incompatible pair is too close compared to some margin.

As illustrated in Figure 4, the embedding of items in the style space are vectors of dimension 256. Following Veit et al. (2015), we measure the compatibility between two items using L_2 norm. The architecture of the CNN in Figure 3 is based on one of the most successful network architectures, GoogLeNet (Szegedy et al., 2015), augmented with a 256-dimension fully connected layer.

3.3.1.1.2 Distinctiveness

We measure the distinctiveness of a clothing item by calculating how visually different the item is from all the other items in the same category. Specifically, we use the embedding in the style space to represent each clothing item's style and calculate the average style for items posted in the past three months (a season), to account for the fact that one style could be distinctive this season but may not be distinctive later on. The distinctiveness of one item is measured by L_2 norm between its style embedding and the average style. This is in a similar spirit to the creativity concept from Toubia and Netzer (2017).

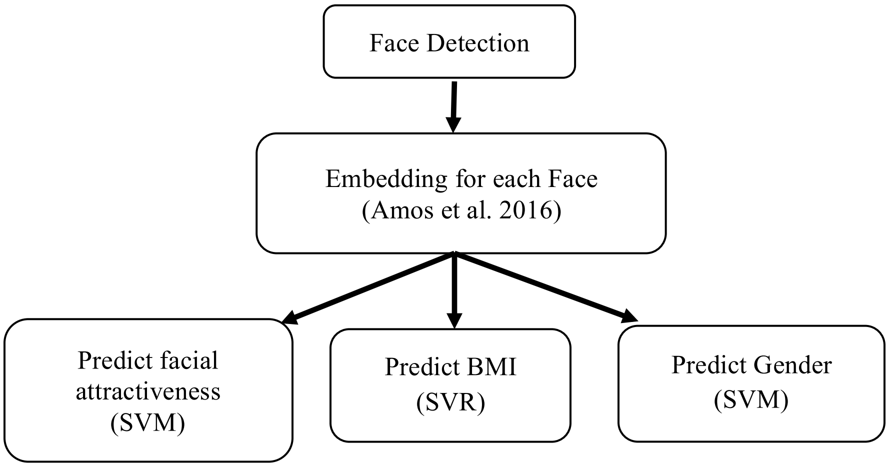
Figure 4. Illustration of Compatibility Feature Extraction



3.3.1.2 User Appearance

For the model styles, we examine the face and body features, specifically the facial attractiveness and body mass index.

Figure 5. Steps of Model Styles Extraction



3.3.1.2.1 Facial attractiveness

The first step is to crop the face and get a vector representation of the face. We follow the deep neural network implementation by the Open Face project (Amos et al., 2016). This architecture was trained for face recognition, providing a 128-dimensional intermediate layer that represents a low dimensional embedding of any face image.

The first step is to get the low dimensional features generated using the deep neural network implementation. Then we need to train a supervised learning model to predict attractiveness. Our training data consist of three thousand images with attractiveness scored on a 1 to 7 scale, where 1 means the face is the least attractive and 7 represents the highest value of attractiveness. Each image is labeled by five Amazon mechanical turkers. We take the average of the five ratings for each image as its final rating. Given the continuous nature of the resulting rating, we train a Support Vector Regression (SVR) model that learns the relationship between the 128-dimensional image features and the attractiveness rating. The model achieves high prediction accuracy with a mean absolute error of 0.66 on the test sample.

3.3.1.2.2 Body feature

We measure BMI to capture the users' body feature. The training data contains 4206 images of faces with true BMI information, made available by Kocabey et al. (2017). These images are collected from Reddit posts linking to the imgur.com service. With the training data, we first crop the faces and get the embedding of 128 dimensions. Then, with the face embedding as the input and BMI (ranging from 11.5 to 50.8) as the output, we again train an SVR model that learns the relationship between the face image feature and the BMI. Eventually, we can predict BMI for a given fashion look, according to the face detected from the fashion look. The model has good performance with a mean absolute error of 2.45.

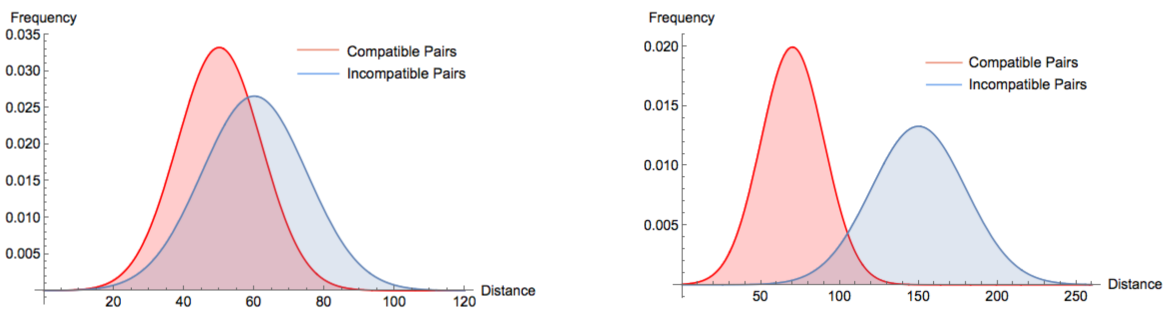
3.3.1.2.3 Gender information enhancement

The gender information for approximately 50% bloggers is not shown on the website. As we also want to know the gender effect, we employ SVM to predict gender from the cropped face images. As a result, we get gender information for 97% bloggers in the whole sample, with 92.89% accuracy on prediction.

3.3.1.3 Results of Feature Extraction

The objective of the Siamese CNN is to project compatible pairs close together and incompatible pairs far away. Figure 6 plots the distribution of distances for compatible and incompatible pairs for both before and after training for transfer learning. The plots show that the fine-tuned neural network separates the two categories with a greater margin and indicates that the network learned to separate compatible from non-compatible clothing items.

Figure 6. Distribution of Distances for Compatible and Incompatible Pairs



(a) Before Training (b) After Training

Table 3. Performance of User Features Extraction

	Facial Attractiveness	BMI	Gender
MAE	0.66	2.45	-
Accuracy	-	-	92.89%

Table 3 reports the performance of the feature extraction tasks on facial attractiveness, BMI, and gender. For SVR task (extracting face attractiveness and BMI), the commonly used performance measure—mean absolute error (MAE)—is reported for the hold-out sample. Accuracy is reported for the binary classification tasks.

In Table 4, we show both high and low score examples of the extracted features, resulting from our trained learning models.

Table 5 shows the summary statistics for the style features we extracted from the images.

Table 4. Example Photos for the Extracted Style Features

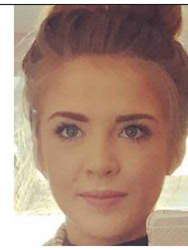
	Low Score		High Score	
Facial Attractiveness				
Body Feature (BMI)				
Compatibility				
Distinctiveness				

Table 5. Summary Statistics for the Extracted Style Features

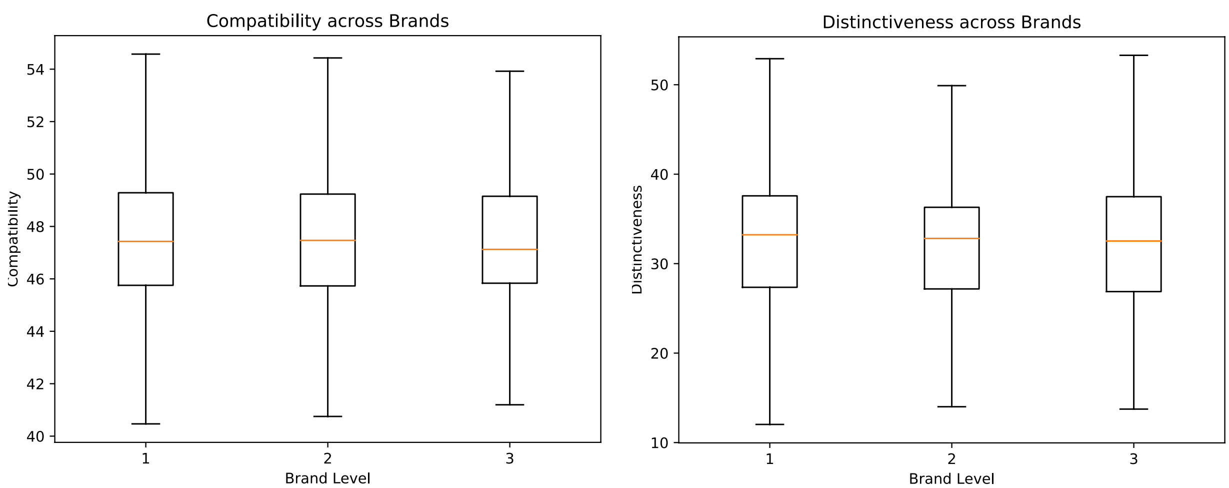
Variable	Mean	Std. Dev.	Min	Max
Facial Attractiveness	4.8201	0.6121	0.3348	7.1819
Body BMI	27.2535	3.6385	11.4966	50.7738
Compatibility	47.4943	12.4622	1.7325	257.1844
Distinctiveness	33.6818	9.8098	8.1391	124.4247

3.3.2 Exploratory Data Analysis

3.3.2.1 Users' Brand and Style Choices

After extracting the style features, we can examine the distribution of styles of fashion looks across the three brand levels to see if the style options are different for each brand level. The boxplots in Figure 7 and the summary statistics in Table 6 show that low-end brands also have pretty good styles, and the style distribution is not much different across brand levels. Therefore, for those who cannot afford luxury brands, there are always substitutes that can provide high-end styles available at lower prices.

Figure 7. Boxplots of Styles across Brands



We also see from Table 6 that the vast majority of fashion looks contain fast fashion brands. It implies that users tend to use at least one fast fashion item to make the whole ensemble look good. Table 7 shows how users mix and match brands from the three brand categories. For fashion looks that adopt mega couture brands, more than 80% of them also adopt fast fashion

Table 6. Summary Statistics at Brand-level

	#item	face	BMI	age	gender	compatibility	distinctiveness	#likes
Fast Fashion (High Street) Brands (Adopted in 99.79% looks)								
mean	1.5652	4.8174	27.2595	24.6837	1.1875	47.4811	33.7177	30.3719
Std.	0.8540	0.6132	3.6407	5.2284	0.3903	12.4851	9.8417	59.7599
Designer Brands (Adopted in 3.05% looks)								
mean	1.0961	4.8533	27.1894	24.8999	1.2192	47.3838	32.9807	57.9064
Std.	0.2948	0.5515	3.5599	3.4180	0.4139	12.5706	9.1262	106.7643
Mega Couture Brands (Adopted in 2.00% looks)								
mean	1.1464	4.9146	26.9429	26.3511	1.1078	48.3523	33.0009	55.9045
Std.	0.3538	0.6242	3.6141	5.2377	0.3103	11.3491	9.1920	89.4968

brands. For those using designer brands, more than 86% of them mix with fast fashion brands. The data evidence shows the economic significance and importance of consumers' mixing and matching behavior in fashion choices.

Table 7. Mix-and-match in Fashion Looks Using Higher-end Brands

(a) For looks adopted mega couture brands

	Only Mega Couture	Mega Couture & Fast Fashion	Mega Couture & Designer	All three
Percentage	4.6%	80.15%	0.37%	14.89%

(b) For looks adopted designer brands

	Only Designer	Designer & Fast Fashion	Designer & Mega Couture	All three
Percentage	3.58%	86.40%	0.24%	9.77%

3.3.2.2 The Impact of Popularity on Choices and Inter-temporal Tradeoff

To understand how consumers choose brand and style, we need to account for a factor that may strongly affect consumers' fashion decision in our research context. That is, users are concerned about popularity or peers' likes. When a user decides whether and what to post, her level of popularity plays a big role. First, being popular can help reduce future cost, given the fact that many popular fashion bloggers are hired by fashion companies and paid to post instead of paying for what they wear. Second, the incentives for posting on social media (Lee et al., 2015) are typically social interaction and self expression, a user may derive higher utility from a given post when she is more popular and has more people watching or following. Therefore, a blogger may be strategic and forward-looking, in the sense that posting to attract others' likes to build up popularity today can help improve future utility because there would be a larger audience for future posts, as well as lower cost for future posts. In other words, when making a post decision, a blogger considers not only the current period's utility, but also the future utility. Even though posting may be worse than not posting in the current period, she may still post because it builds up popularity and the blogger can gain much more in the future.

We have explained that the discrete choice of posting is dynamic as popularity affects one's utility of posting. Moreover, a user also faces inter-temporal tradeoff when making the brand and style choices. Because the brand and style choices affect peer likes a post can attract, but the best choices for attracting likes and building up popularity may not be the choices to satisfy one's own per-period intrinsic preference. When one is not that popular, she may focus on attracting peer likes and less on self-intrinsic taste. In comparison, a popular blogger can make style and brand choices subject to less financial constraint and focus more on expressing her intrinsic fashion tastes.

An ideal measure of popularity is the number of followers. As we do not observe the number of followers for each user across time, we use the cumulative sum of likes ("SumLike," henceforth) a person has gotten from previous posts as a good proxy for his/her popularity. Figure 8 shows the positive linear relationship between SumLike and the number of followers at a snapshot time Aug 1st, 2017. The correlation between SumLike and the number of followers by the time Aug 2017 is 0.85, while that between the average number of likes and number of followers is 0.55. Therefore, we choose SumLike as a proxy measure for one's popularity or, more generally, the exposure one may get when posting a new fashion look.

Figure 8. Followers versus Likes

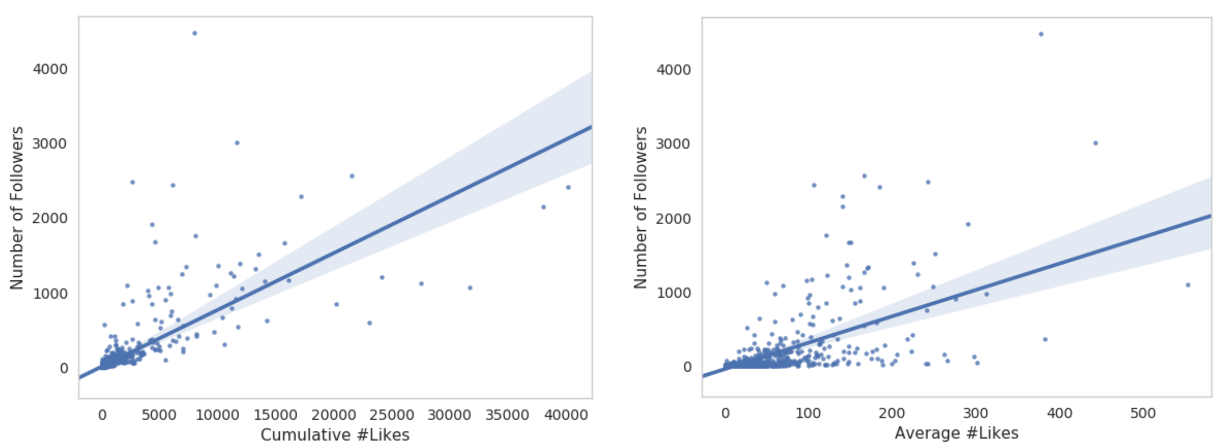


Table 8. Popularity Affects Decisions on Fashion Look

VARIABLES	DV: Brand 1	DV: Brand 2	DV: Brand 3	DV: Compatibility	DV: Distinctiveness
Popularity	-0.0115*	-0.00101	0.00278**	0.292***	0.0331
	(0.00609)	(0.00135)	(0.00109)	(0.0628)	(0.0449)
Observations	18,957	18,957	18,957	18,957	18,957
R-squared	0.639	0.272	0.391	0.192	0.219

Note: Fixed effects at individual level; Robust standard errors, clustered at individual level.
*p<0.01;**p<0.001;***p<0.0001

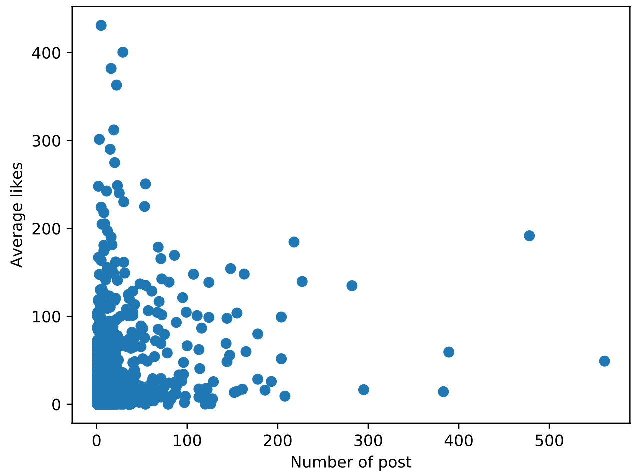
In Table 8, we show the results of five separate regressions to see how popularity up to period $t-1$ affects one's choice of brands and styles in period t. We can see that when bloggers become more popular, they will post more mega couture brands but fewer fast fashion brands, echoing the cost decreasing effect of being popular. Moreover, they further improve the style, implying that the marginal return of improving style is higher when more people are watching.

3.3.2.3 Incentives of Posting on Social Media

People post on social media for different reasons: self-expression, social interaction, archiving, escapism, and so on (Lee et al., 2015). Below is the scatter plot of the number of posts and average likes for each user's posts. We can see that those who post more do not necessarily get more likes than users who post less. Some users' posts on average attract more than four hundred likes but the users post only around ten times, whereas other users' posts get few likes but the users post hundreds of times. Therefore, we hypothesize that the users' utility of posting is not driven only by others' likes or opinions; they may also derive utility from other channels. For example, an individual might enjoy expressing herself through posting a fashion picture, or she wants to attract those who have the same tastes. In these cases, getting likes is not the primary goal.

As a result, when modeling bloggers' decision processes, we adopt a general utility functional form, which captures not only the impact of others' likes but also bloggers' intrinsic utility derived from the brand and style choices.

Figure 9. Scatterplot of Posts and Likes



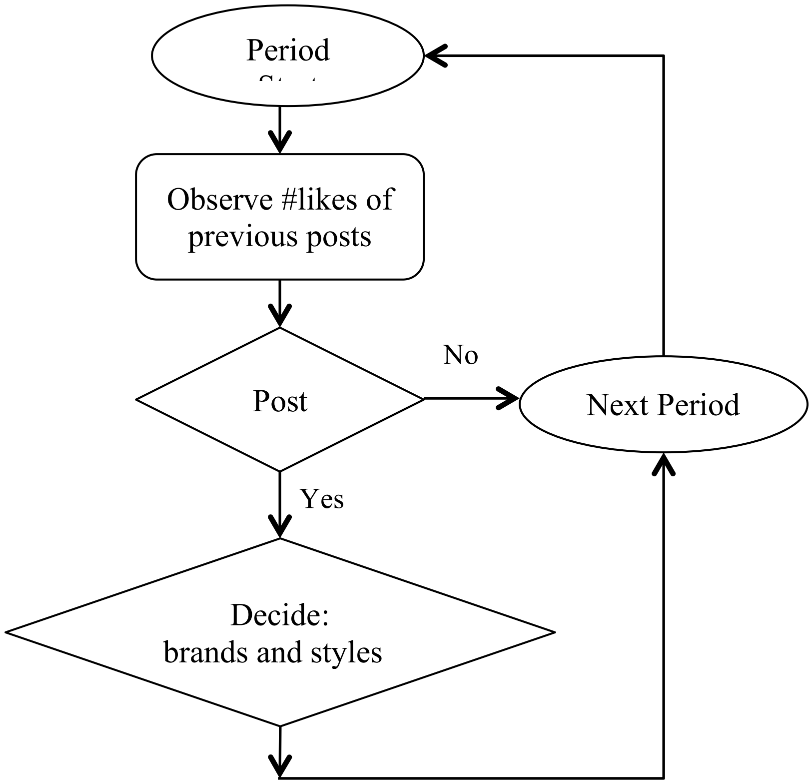
However, with the above exploratory analysis, it is still not clear how consumers value brands and styles, and how substitutable brands and styles are. Can good styles from low-end brands compensate for the utility loss from not being able to get the high-end brand with relatively worse styles? We proceed to answer this question by modeling blogger decisions at the micro-level in the next section.

3.4 Model

The timing of the events is illustrated in Figure 10. In each period:

1. The blogger observes the popularity she has built up with her previous posts (if there are any).
2. She decides whether to post a fashion look. If she posts, she simultaneously chooses the brand and style. If not, she goes to the next period.
3. The number of peer likes for the new post (if there are any) realizes. The next period comes.

Figure 10. Timing of Events



3.4.1 The Basic Model

Let St_{it}, Br_{it} denote the decision on style and brand. P_{it} is the binary decision of posting. The per-period utility of blogger i in period t follows a nested constant elasticity of substitution (CES) format (Solow, 1956):

$$u_{it}(St_{it}, Br_{it}, P_{it}|S_{it}, \theta_i) = F_{it}(S_{it}, \theta_i) \cdot R_{it}(St_{it}, Br_{it}|\theta_i) \cdot \mathbf{1}\{P_{it} = 1\} + \epsilon_{it}$$

where $R_{it}(St_{it}, Br_{it}|\theta_i)$ captures the utility gain from the fashion look attributed to the brand Br_{it} and style St_{it}.

$$R_{it}(St_{it}, Br_{it}|\theta_i) = [\alpha_{i1} {Br_{it}}^{\rho_{\mathrm{i1}}} + (1 - \alpha_{i1}) {St_{it}}^{\rho_{\mathrm{i1}}}]^{\frac{1}{\rho_{\mathrm{i1}}}},$$

where the elasticity of substitution between style and brand is $r_{i1} = \frac{1}{1-\rho_{\mathrm{i1}}}$, and α_{i1} is the share parameter. In our study, we regard the brand value as the incremental value associated with the brand name that is not related to any style attribute. For example, apart from any style features, people choose luxury brands because they work as a social label and provide hedonic rewards.

$F_{it}(\cdot)$ captures the valuation of a post. It measures the effect of popularity and individual fixed effect on the utility, apart from the brand and style choices.

$$F_{it}(S_{it}, \theta_i) = \eta_{i0} + \eta_{i1} \ln\left(1 + SumLike_{i,t-1}\right).$$

$SumLike_{i,t-1}$ is the cumulative number of likes user i got from the fashion looks that she posted in the past $t-1$ periods. In the per-period utility function, ϵ_{it} is the brand-and-style choice specific random error, following a Type-1 extreme value distribution.

Brand choice

The brand choices are discrete. The blogger makes brand choice B_{it} from categories of fast fashion (level 1), designer (level 2), and mega couture (level 3), respectively denoted by $l \in \{1,2,3\}$. For each brand level, the blogger decides how many items to include in a post. The brand choice for a fashion post is characterized by the following linear function.[44]

$$Br_{it} = exp\left(\sum_{l=1}^{3} \gamma_{i,l} x_{it,l} + \sum_{l \neq k} \gamma_{i,lk} x_{it,l} x_{it,k}\right), \quad (2)$$

where $x_{it,l}$ denotes the number of clothing items of brand category l, in the fashion look posted by blogger i in period t. γ_{il} captures the utility gain of choosing an additional item of brand l. $\gamma_{i,lk}$ measures the utility gain of matching items from brand category l and k.

Style choice

The bloggers make style choices. The style, as a factor of the whole fashion look, is incorporated as a sub-nest of the CES utility function.

$$St_{it} = \left[\alpha_{i2} f_{it,1}^{\rho_{i2}} + (1-\alpha_{i2}) f_{it,2}^{\rho_{i2}}\right]^{\frac{1}{\rho_{i2}}}$$

where α_{i2} is the share parameter, and $f_{it,1}, f_{it,2}$ denote the choices of compatibility and distinctiveness. The elasticity of substitution between the two style features is $r_{i2} = \frac{1}{1-\rho_{i2}}$.

Budget and cost

The blogger's decisions are subject to a budget constraint y_i. Specifically,

$$\sum_{j=1}^{k} t_{ij} f_{it,j} + \sum_{l=1}^{3} t_{ix,l} x_{it,l} \quad \leq y_i.$$

The cost of purchasing a clothing item is allowed to change with $Like_{t-1}$.

$$t_{ix,l} = \frac{\tilde{t}_{i,l}}{1 + \delta_i \cdot \ln\left(1 + SumLike_{i,t-1}\right)},$$

[44] We apply exp(.) to make sure the brand choice measure is positive so that it is a valid input of the CES function. If there is no brand information for a tagged item, we assume others' belief is level 1; otherwise, the user would disclose the brand information. This is consistent with the data fact that the majority of brands are level 1 demonstrated in Appendix, and conforms to consumer rationality, given higher-end brands attract more likes.

where $l \in \{1,2,3\}$, and $\tilde{t}_{i,l}$ is the baseline cost for obtaining an item with brand-level l. δ_i measures the decreasing effect of the cumulative number of likes on the cost of brand choices. Bloggers may incur less cost if they are valuable to the fashion companies, as influencers can help market their products to the public. The more followers (measured by $SumLike_{i,t-1}$) a user has, the more valuable she is to the fashion company, because her posts will influence a broader audience who potentially can become customers. This is demonstrated by the fact that many fashion bloggers are paid millions every year by fashion companies.[45]

For identification purposes, we normalize the base cost for a fast fashion brand to 1—that is, $\tilde{t}_{i,1} = 1$. The budget constraint for blogger i, y_i, measures the highest cost a blogger is willing to pay for the fashion consumption, and it is assumed fixed over weeks, as the time and pecuniary resources allocated to other regular activities (e.g., working, entertainment) typically do not change much over weeks. In estimation, the budget constraint is calculated by the highest ever cost since the user started posting.

3.4.2 State Variables

The state variables are $S_{it} = \{face_i, body_i, age_i, gender_i, SumLike_{i,t-1}, \epsilon_{it}\}$, where $SumLike_{i,t-1}$ is time varying while others are fixed for the same user. With $\boldsymbol{f_{it}} = (f_{it,1}, f_{it,2})$ and $\boldsymbol{x_{it}} = (x_{it,1}, x_{it,2}, x_{it,3})$, the state transition follows

$$Like_{it} = \widehat{g}\,(SumLike_{i,t-1}, face_i, body_i, age_i, gender_i, \boldsymbol{f_{it}}, \boldsymbol{x_{it}}) + \zeta_{it}$$

where $\widehat{g}\,(\cdot)$ is estimated by a linear regression as the first step, and $\zeta_t \sim N(0, \widehat{\sigma})$. In our dataset, we observe $\hat{S}_{it} = \{face_i, body_i, age_i, gender_i, SumLike_{i,t-1}\}$. The state transition regression results are shown in section 6.

3.4.3 Inter-temporal Tradeoff

Each user decides on an infinite sequence of decision rules $\{f_{it}, x_{it}, post_{it}\}_{t=1}^{\infty}$ to maximize the expected discounted utility. Substituting the brand and style choices (equation (2) and (3)) to the per-period utility function (equation (1)), we have

$$\max_{\{f_{it}, x_{it}, P_{it}\}_{t=0}^{\infty}} E_{\{S_{it}\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta^t \, U_t(\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it} | S_{it}, \theta_i) | x_{i0}, f_{i0}, P_{i0}, S_{i0} \right\},$$

where

[45] http://www.blogingrace.com/highest-paid-fashion-bloggers/

$$
\begin{aligned}
&U_t(\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it}|S_{it}, \theta_i) \\
&\qquad = [\eta_{i0} + \eta_{i1}\ln(1 + SumLike_{t-1})] \\
&\qquad \cdot \left\{ \alpha_i \left[exp\left(\sum_{l=1}^{3} \gamma_{i,l} x_{it,l} + \sum_{l \neq k} \gamma_{i,lk} x_{it,l} x_{it,k} \right) \right]^{\rho_{i1}} \right. \\
&\qquad \left. + (1-\alpha_i)\left[\beta_i f_{1it}^{\rho_{i2}} + (1-\beta_i) f_{2it}^{\rho_{i2}}\right]^{\frac{\rho_{i1}}{\rho_{i2}}} \right\}^{\frac{1}{\rho_{i1}}} \cdot \mathbf{1}\{P_{it} = 1\} + \epsilon_{it}
\end{aligned}
$$

Let $V(S_{it})$ denote the value function:

$$
\begin{aligned}
V(S_{it}) = &\max_{\{\boldsymbol{f_{i\tau}}, \mathbf{x_{i\tau}}, P_{i\tau}\}_{\tau=t}^{\infty}} E_{\{S_{i\tau}\}_{\tau=t}^{\infty}} \left\{ U_t(\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it}|S_{it}, \theta_i) \right. \\
&\left. + \sum_{\tau=t+1}^{\infty} \beta^{\tau-t}\, U_\tau(\boldsymbol{f_{i\tau}}, \boldsymbol{x_{i\tau}}, P_{i\tau}|S_{i\tau}, \theta_i) | x_{it}, f_{it}, P_{it}, S_{it} \right\}
\end{aligned}
$$

The Bellman Equation (Bellman, 1957) for the dynamic optimization problem is expressed as follows:

$$
V(S_{it}) = \max_{f_{it}, x_{it}, \mathrm{P}_{it}} E_{S_{it+1}} \left\{ U_t\left(\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it}|S_{it}, \theta_i\right) + \beta V(S_{it+1}) | \boldsymbol{x_{it}}, \boldsymbol{f_{it}}, P_{it}, S_{it} \right\}
$$

All the decisions, that is, $\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it}$, are dynamic in the sense that the current period's decisions affect the next period state through $SumLike_{t-1}$, which further affects the cost and utility of the user's brand and style choices. In other words, the users face inter-temporal tradeoffs regarding the utility derived from the post today versus its impact on the utility of posting tomorrow.

3.4.4 Heterogeneity

In the fashion sharing community, users may have different responses to others' opinions. For example, some users care a lot about others' opinion (i.e., #likes), whereas some care only about expressing themselves rather than attracting #likes. Similarly, some users base their utility heavily on the brand levels, whereas others care more about the clothing styles. Therefore, we employ a hierarchical Bayesian framework (Rossi et al., 2005) to account for heterogeneity. All structural parameters $\theta_i \in \boldsymbol{\Theta_i} = \{\alpha_{i1}, \alpha_{i2}, \boldsymbol{\rho_i}, \boldsymbol{\eta_i}, \boldsymbol{t_i}, \boldsymbol{\gamma_i}, \delta_i\}$ have random coefficient specification.[46] The prior distribution is normal for $\boldsymbol{\eta_i}, \boldsymbol{\gamma_i}$, and log normal for $\alpha_{i1}, \alpha_{i2}, \boldsymbol{r_i}, \boldsymbol{t_i}, \delta_i$, where $\boldsymbol{r_i} = \frac{1}{1-\boldsymbol{\rho_i}}$. The prior distribution is specified as $N(\mu_\theta, \sigma_\theta^2)$. We use diffuse hyper-prior distribution for all parameters.

---

[46] The parameters in bold are vectors with subscription 1,2, etc. as specified in the model.

3.5 Identification and Estimation

3.5.1 Identification

In our model, the unknown parameters include those in the state transition process ($\vec{b}$) and the primitives in the utility function ($\boldsymbol{\Theta_i}$). We briefly explain the identification of the key parameters.

First of all, the transition function parameters can be identified by the variation in the number of likes $Like_{i,t}$ and the corresponding fashion brands, style features, and blogger demographics (i.e., $SumLike_{i,t-1}, face_i, body_i, age_i, gender_i, \boldsymbol{f_{it}}, \boldsymbol{x_{it}}$). For the discount factor β, as acknowledged in the literature (e.g., Rust, 1987), it cannot be separately identified from the utility parameter, that is, the individual-fixed effect in our context. Therefore, following the conventional approach, we fixed the weekly discount at 0.998 to stay consistent with the literature (Hartmann & Nair, 2010; Liu et al., 2018).

Table 9. Summary of the Parameters

Notation	Explanation
α_{i1}	Share parameter for brand and style.
α_{i2}	Share parameter for different style features.
$\boldsymbol{\rho_i}$	Elasticity of substitution, $\boldsymbol{\rho_i} = \{\rho_{i1}, \rho_{i2}\}$, for brand versus style, and between style features.
$\boldsymbol{\eta_i}$	Governing the efficiency or productivity. $\boldsymbol{\eta_i} = \{\eta_{i0}, \eta_{i1}\}$, where η_{i0} captures individual fixed-effect, η_{i1} measures the effect from the cumulative likes.
$\boldsymbol{t_i}$	Cost parameters for brand and style choices. $\boldsymbol{t_i} = \{t_{i1}, t_{i2}, t_{ix2}, t_{ix3}\}$.
$\boldsymbol{\gamma_i}$	Utility gain from an item of a certain brand level, $\boldsymbol{\gamma_i} = \{\gamma_{\mathrm{i1}}, \gamma_{\mathrm{i2}}, \gamma_{\mathrm{i3}}, \gamma_{\mathrm{i,12}}, \gamma_{\mathrm{i,13}}, \gamma_{\mathrm{i,23}}\}$.
δ_i	The costing decreasing effect from cumulative #likes (as a proxy for #followers).

The elasticity of substitution between brand and style can be identified by the variation in the brand (Br_{it}) and style choices (St_{it}) across time. Similarly, the elasticity of substitution between compatibility and distinctiveness can be identified by the variation in the two style choices across periods. The brand and style choices help us back out the underlying cost of each blogger. We can separately identify the effect of $SumLike$ on the utility (η_{i1}) and on the cost (δ_i), because the cost decreasing effect is only on the brand choices, whereas η_{i1} affects both the brand and style choices in the same way. So, the different evolving patterns of brand and style choices can help us identify η_{i1} and δ_i.

In summary, the structural parameters to be estimated are $\boldsymbol{\Theta_i} = \{\alpha_{i1}, \alpha_{i2}, \boldsymbol{\rho_i}, \boldsymbol{\eta_i}, \boldsymbol{t_i}, \boldsymbol{\gamma_i}, \delta_i\}$.

3.5.2 Likelihood

The likelihood function is

$$Likelihood = L\left(\left\{\left\{\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it} \middle| \hat{S}_{\mathrm{it}}, \theta_i\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) L\left(\left\{\widehat{g}\left\{\hat{S}_{it} \middle| \hat{S}_{it-1}, \boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{\mathrm{it}}\right\}_{\mathrm{t}=1}^{T}\right\}_{i=1}^{I}\right) L\left(\left\{\hat{S}_{\mathrm{i}0}\right\}_{i=1}^{I}\right)$$

where $\hat{S}_{it}$ includes all the observable states. According to the above likelihood function, the likelihood for the state transition process and that for the optimal choices can be separately estimated. Our data cover the entire history of activities for each individual, and everyone starts with $Like_{\mathrm{i}0} = 0$. As the first step, we estimate the state transition process, $\widehat{g}\,(\cdot)$, with a linear regression. Then we maximize the likelihood of the optimal choices:

$$L\left(\left\{\left\{\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it} \middle| \hat{S}_{it}, \theta_i\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) = \prod_{i=1}^{I}\prod_{t=1}^{T} L\left\{\boldsymbol{f_{it}}, \boldsymbol{x}_{\mathrm{it}}, P_{it} \middle| \hat{S}_{it}, \theta_i\right\},$$

where the brand choices $\boldsymbol{x_{it}}$ are discrete and style choices $\boldsymbol{f_{it}}$ are continuous.

The likelihood for each choice $\{\boldsymbol{f_{it}}, \boldsymbol{x}_{\mathbf{it}}, P_{it}\}$, consisting of both discrete and continuous choices, can be calculated through a discrete way. Note that for each combination of choice, $\{\boldsymbol{f_{it}}, \boldsymbol{x}_{\mathbf{it}}, P_{it}\}$, the predicted number of likes can be obtained with the estimated transition regression model—that is, $\widehat{Like}_{it} = \widehat{g}\,(Like_{i,t-1}, face_i, body_i, age_i, gender_i, \boldsymbol{f_{it}}, \boldsymbol{x}_{\mathbf{it}}) \cdot \mathbf{1}\{P_{it} = 1\}$. Then, we can first eliminate all style choices that are strictly dominated. Specifically, the assumption of blogger rationality implies a unique choice set $\{\boldsymbol{f_{it}}, \boldsymbol{x}_{\mathbf{it}}, P_{it}\}$ corresponding to the set $\{\widehat{Like}_{it}, \boldsymbol{x}_{\mathbf{it}}, P_{it}\}$. With the CES utility functional form, there exist unique closed-form solutions for $\boldsymbol{f_{it}^*}$ given $\{\widehat{Like}_{it}, \boldsymbol{x}_{\mathbf{it}}, P_{it}\}$. To put it in math, given $\{\boldsymbol{x}_{\mathbf{it}}, P_{it}\}$, the optimal choices of $\boldsymbol{f_{it}^*}$ are unique for each $\widehat{Like}_{it}$, obtained by solving the following maximization problem:

$$\max_{\{f_{ijt}\}_{j=1}^{\mathrm{k}}} \left[\beta_i f_{1\mathrm{it}}^{\rho_{i2}} + (1-\beta_i) f_{2\mathrm{it}}^{\rho_{i2}}\right]^{\frac{1}{\rho_{i2}}}$$

subject to $\widehat{g}\left\{\boldsymbol{f_{it}} \middle| \boldsymbol{x}_{\mathbf{it}}, P_{it}, \hat{S}_{it}\right\} = \widehat{Like}_{it}$ and $\sum_{j=1}^{k} t_j\, f_{ijt} + \sum_{l=1}^{3} t_{xl} x_{it,l} \ \leq y_i$.

Therefore, styles choices that satisfy $\boldsymbol{f'_{it}} \neq \boldsymbol{f^*_{it}}$ but lead to the same $\widehat{Like}_{it}$ are strictly dominated, thus will never be chosen. With this observation, we can further relieve the computation burden by acting on $\widehat{Like}_{it}$ instead of multiple choices $\boldsymbol{f_{it}}$. Please see the appendix for details about how we transform the continuous choice space for $\boldsymbol{f_{it}}$ into a discrete action space on $\widehat{Like}_{it}$ while reserving the continuous nature of the style choices.

The likelihood of the optimal choices can be expressed as

$L\left(\left\{\left\{\boldsymbol{f_{it}}, \boldsymbol{x_{it}}, P_{it} \middle| \hat{S}_{it}, \theta_i\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) = \prod_{i=1}^{I} \prod_{t=1}^{T} L\left\{\boldsymbol{f_{it}}, \boldsymbol{x}_{\mathbf{it}}, P_{it} \middle| \hat{S}_{it}, \theta_i\right\} =$
$\prod_{i=1}^{I} \prod_{t=1}^{T} Pr\left\{\widehat{Like}_{it}, \boldsymbol{x}_{\mathbf{it}}, P_{it} \middle| \hat{S}_{it}, \theta_i\right\}.$

With Type-1 extreme value distribution for the random error, the choice probability can be written as:

$$Pr\left\{\widehat{Like}_{it,\mathrm{n}}, \boldsymbol{x}_{\mathbf{it},\boldsymbol{n}}, P_{it,n} \middle| \hat{S}_{it}, \theta_i\right\} = \frac{\exp\left\{?\left(\widehat{Like}_{it,\mathrm{n}}, \boldsymbol{x}_{\mathbf{it},\boldsymbol{n}}, P_{it,n} | \hat{S}_{it}, \theta_i\right)\right\}}{\sum_{n=1}^{N} \exp\left\{v\left(\widehat{Like}_{it,\mathrm{n}}, \boldsymbol{x}_{\mathbf{it},\boldsymbol{n}}, P_{it,n} | \hat{S}_{it}, \theta_i\right)\right\}}$$

where

$v(\widehat{Like}_{it,\mathrm{n}}, \boldsymbol{x}_{\mathbf{it},\boldsymbol{n}}, P_{it,n} | \widehat{S}_{it}) = U_t\left(\boldsymbol{f}_{\boldsymbol{it},\boldsymbol{n}}, \boldsymbol{x}_{\mathbf{it},?}, P_{it,n} | S_{\mathrm{it}}\right) - \epsilon_{it,n},$

and $\boldsymbol{f_{it,n}}$ are observed from the data or backed out from $\left\{\widehat{Like}_{it,\mathrm{n}}, \boldsymbol{x}_{\mathbf{it},\boldsymbol{n}}\right\}$ if there is no post in that period. N is the total number of discrete choices. According to the data, in a fashion look, there are up to three items tagged as brand level 1, while there are up to two items tagged as brand level 2 (designer) and 3 (couture).[47] For fashion looks without brand tags, we assume audiences' belief about the brand level is fast fashion (level 1), which is consistent with the majority observation and rational behavior of tagging.[48] Therefore, there are $3 \times 2 \times 2 - 1 = 35$ brand choices if one decides to post in our context. The choice of target #likes can be from 0 to the largest #likes the bloggers have gotten on the website, which is 1747. Together with the choice of not posting, the bloggers have in total $N = 35 \times M + 1$ discrete choices. For ease of computation, we can shrink the choice space for an individual according to the historical #likes she has gotten, plus some deviation based on $\widehat{\sigma}$.

3.5.3 Estimation Methods

To estimate the infinite horizon dynamic structural model, we explore several methods, including the conditional choice probability estimation (CCP) (Hotz & Miller, 1993; Aguirregabiria & Mira, 2007; Arcidiacono & Miller, 2011), the simulated method of moments (SMM) (Pakes & Pollard, 1989; McFadden, 1989), and the Bayesian estimation method (Imai, Jain, & Ching, 2009 (IJC)). SMM matches data moments with simulated moments, but it requires fully solving the dynamic optimization problem and is thus computationally very costly; moreover, it cannot capture the rich heterogeneous responses across different individuals. The CCP methods improve on the computational efficiency but can recover only very limited heterogeneity. The IJC method serves our goal to capture rich individual responses, and the

---

[47] There are only fewer than 1% posts tagging more than the upper limits of brands; therefore, we bound the choices below 3, 2, 2 respectively for each level of brand. Though there are cases where people wear the same clothing item in multiple posts, they do not tag the same item, and there must some other new items across different posts.

[48] If the items are of higher-level brands, the consumer would tag them given the belief that no-tagged brands are level 1, because the fixed-effect regression indicates that a higher-level brand contributes positively to #likes.

computational burden is also alleviated because it requires evaluating the value function only once in each iteration.

However, IJC is designed for discrete choice models, but our model includes both discrete and continuous choices. With the modification for the dynamic choice problem explained in section 5.2 (please see more details in Appendix D), we can apply the IJC while reserving the continuous nature of the style choices.

3.6 Results

3.6.1 Model Comparison

We compare our model with the benchmark model without the forward looking and direct utility derived from followers. When bloggers are myopic, they do not consider the impact of today's choice on tomorrow's state, that is, number of followers. When there is no direct utility gain from followers, the only effect of followers is through decreasing cost. The utility gain from followers adds more to the model fitting than forward looking. All four measures show that the proposed model outperforms the benchmarks significantly.

Table 10. The Proposed Model vs. Alternative Models

	No forward-looking	No utility from followers	Proposed
Hit Rate: Post	0.564	0.532	0.877
Hit Rate: Fast Fashion	0.775	0.610	0.922
Hit Rate: Designer	0.589	0.533	0.897
Hit Rate: Mega Couture	0.520	0.421	0.811

3.6.2 Parameter Estimates

Table 11 shows the transition process estimated with an OLS regression. We can see that popularity does have a substantial positive effect on the peer likes for a new fashion post. The number of clothing items has increasingly positive effects with higher brand levels. More attractive faces attract more likes. The average audience likes a lower BMI, younger looks, and male models. We also see a significant positive effect of compatibility. The distinctiveness does not show a significant impact on attracting likes. The negative effect of the interaction term between high-street brands and mega couture brands shows that people do not respond favorably to outfits that match low and high-end brands together. However, the positive effect of including either a fast fashion brand or a mega couture brand dominates the negative interaction effect.

Table 12 reports the results for structural parameters. The estimation converges with 7000 iterations. We ran 10,000 iterations and report the results using the last 2000 iterations after burn-in.[49] Figure 11 shows the histogram of some structural parameters across individual bloggers. The mean elasticity of substitution is $r_{1i} = 1/(1-\rho_{1i}) = 1.24$ and implies that styles and brands are quite substitutable for most bloggers. In other words, they find it easy to substitute a high-end brand with good style to derive the same utility. As shown in Figure 11, 60.11% of the individuals treats style and brands as substitutes. For this group of consumers, high-end brands may need to worry about the copycats' cannibalization effect. For the the rest 39.89% users, they view style and brand as complements, meaning they will not be lured by only good styles, which makes the copycat problem less worrisome.

Table 11. Regression of State Transition

VARIABLES	OLS	Standard Error
Log (1+#Like_t-1)	12.74***	0.147
# Fast fashion (level 1)	8.814***	0.466
# Designer (level 2)	15.72**	7.237
# Mega couture (level 3)	24.18***	7.180
Face	5.768***	0.744
BMI	-0.644***	0.121
Age	-0.485***	0.0819
Gender	2.952***	0.928
Compatibility	0.151***	0.0351
Distinctiveness	0.0116	0.0448
Level 1*Level 2	-0.604	2.812
Level 1*Level 3	-6.937**	2.963
Level 2*Level 3	-1.087	6.935
Constant	-45.87***	7.532
Observations	21,093	
R-squared	0.304	

Note: "#" denotes "the count of"; Robust standard errors; Observations are collapsed at weekly level; *p<0.01, **p<0.001, ***p<0.0001.

The estimates of η_{1i} indicate that most people (about 80%) value popularity or others' attention, but there exists large heterogeneity across individuals, with a standard deviation of 0.3258. There are about 21.92% users have negative value for popularity, probably because they feel uncomfortable or stressful when more people are watching and paying attention to them. This group of users may post for archiving or escapism, which requires less or no other people's attention. From the third plot of Figure 11, we see there is also a large cost decreasing effect of popularity. This is consistent with the fact that influential bloggers are hired by fashion companies to promote their products.

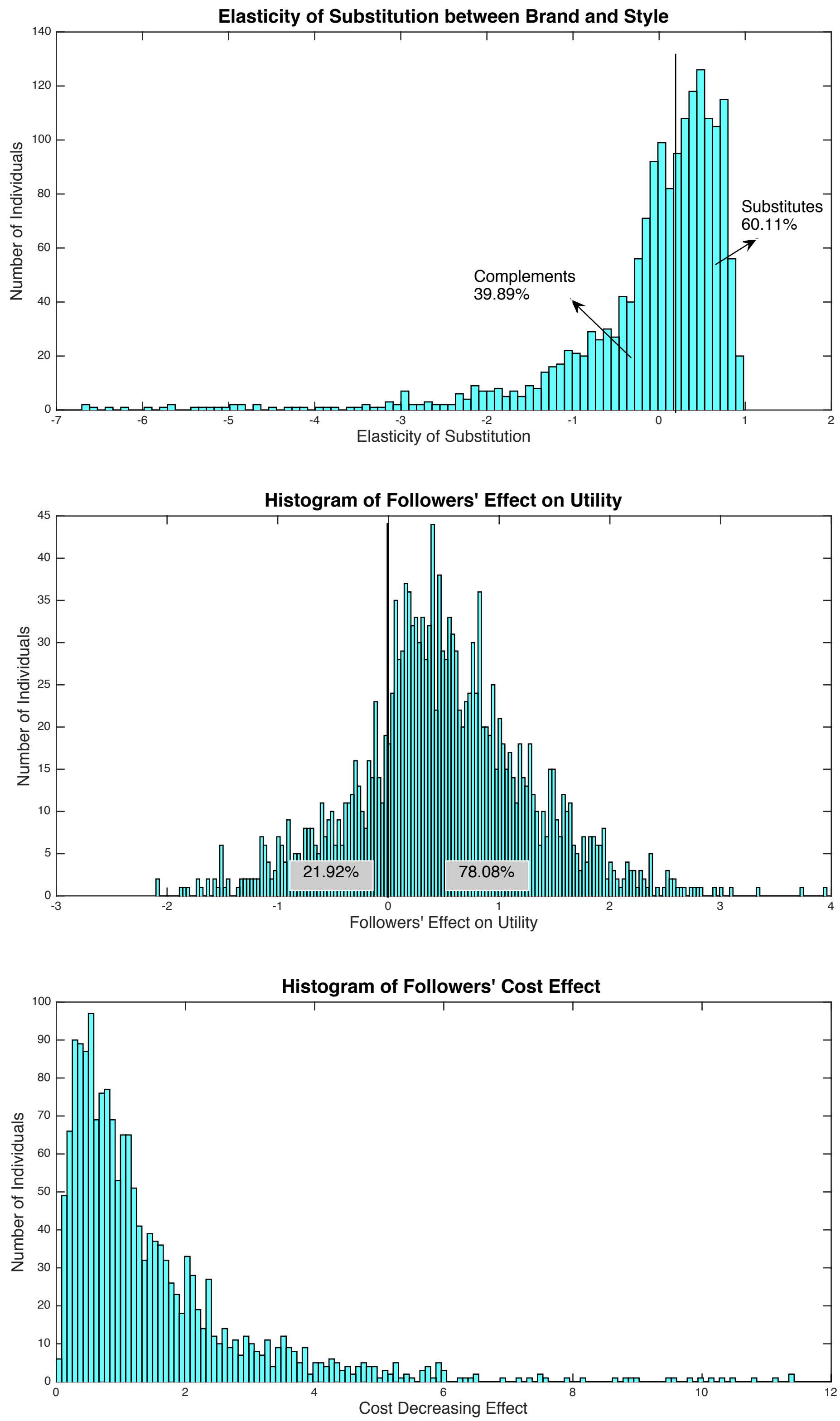
[49] Convergence was visually assessed with plots of the structural parameters. We store 100 past pseudo-value functions.

Table 12. Structural Model Estimation Results

Variable	Interpretation	Mean (μ_θ)	Standard deviation (σ_θ)
α_{i1}	Share parameter	**0.5143**	0.1068
		(0.0059)	(0.0181)
α_{i2}		**0.5812**	0.0971
		(0.0054)	(0.0172)
ρ_{1i}	Elasticity of substitution	**0.1903**	0.2759
		(0.0173)	(0.0135)
ρ_{2i}		**-0.1269**	0.4131
		(0.0221)	(0.0151)
η_{0i}	Fixed effect	**0.1557**	0.4497
		(0.0265)	(0.0205)
η_{1i}	Effect of cumulative #likes	**0.4926**	0.3258
		(0.0181)	(0.0152)
t_{1i}	Cost for style choices	**1.0768**	0.4277
		(0.0263)	(0.0178)
t_{2i}		**0.9369**	0.3713
		(0.0213)	(0.0172)
$t_{x2,i}$	Cost for brand choices	**2.0601**	0.4103
		(0.0224)	(0.0143)
$t_{x3,i}$		**2.9698**	0.8729
		(0.0292)	(0.0177)
γ_{1i}	Gain from brand choices	**1.1805**	0.4763
		(0.0311)	(0.0174)
γ_{2i}		**0.6857**	0.2827
		(0.0161)	(0.0162)
γ_{3i}		**0.7848**	0.2917
		(0.0208)	(0.0178)
$\gamma_{12,i}$		-0.0130	0.3994
		(0.0212)	(0.0177)
$\gamma_{13,i}$		-0.0269	0.4013
		(0.0228)	(0.0176)
$\gamma_{23,i}$		0.0050	0.3757
		(0.0209)	(0.0146)
δ_i	Cost decreasing effect from cumulative #likes.	**0.9902**	0.3821
		(0.0220)	(0.0156)

Note: Standard errors in parentheses.

Figure 11. Distribution of Structural Parameters across Individuals



3.7 Counterfactual Studies

3.7.1 If Fast Fashion Cannot Copy Mega Couture Styles

We wanted to see what would happen to consumers' choices if copyright law provided more protection for fashion designs. In this counterfactual world, the fast fashion companies are prohibited from producing styles similar to the original fashion styles of mega couture brands. As a result, the style choices at the fast fashion level would be more restricted. Specifically, the styles of the high-end brands become relatively more exclusive or distinctive, whereas the styles at the fast fashion level become not distinctive, according to our measurement of distinctiveness. Moreover, as fast fashion companies cannot produce copycat styles in the counterfactual world, if consumers wear both the high-end and low-end styles together, the compatibility would be much lower than before. How would exactly the style options change? For each mega couture item, we first calculate its distance to all other fast fashion items, using the vector representation. We experimented by dropping the fast fashion styles with 10% and 5% smallest distance (largest similarity) and replace the removed ones with the average of the other fast fashion styles. Then we calculate the distinctiveness for the fast fashion styles, and the highest compatibility between a fast fashion and a higher-end brand. We observe the remaining fast fashion styles are below about 25 percentile of the mega couture styles. Therefore, we operationalize the analysis by restricting the distinctiveness of the fast-fashion styles to be bounded below 25 percentiles of the styles from the mega couture brands. We also restrict the compatibility of matching fast fashion with mega couture brands to be within the bottom quartile of other combinations.

The first column of Table 13 shows the results of a hundred simulations. On average, the posting probability drops by 7.92%. Interestingly, not only the fast fashion brands are worse off, but also the adoption of all brands decreases. Why are the high-end brands also worse off when copycats are prohibited? By comparing consumers' choices in the counterfactual world and what they chose before, we found three mechanisms contributing to a lower demand for the mega couture brands.

First, many bloggers combine clothing items across brand levels to make a complete outfit. When the style choice is restricted at low-end brands, consumers are subject to higher financial pressure to buy high-end brands to get a satisfactory ensemble. In other words, this is driven by the consumers who cannot always afford high-end brands for their ensembles and who therefore mix and match both low-end and high-end brands. The counterfactual policy would put this group of consumers unsatisfied with the styles of mixing and matching high-end and low-end brands together, resulting in them buying nothing. Thus, they end up buying fewer clothes and post less. The data facts (Table 7) suggest that about 80% (86%) of fashion looks incorporating a mega couture (designer) brand also include items at the fast-fashion level, further demonstrating the economic significance of the mix-and-match mechanism in the fashion market.

Second, some consumers may not value high-end brands that much. But they can accumulate popularity across time by wearing attractive styles from fast fashion. Once their popularity is high, there are more people following and paying attention to them, then they will derive more value from what they wear, including high-end brands. Therefore, they will be more likely to adopt high-end brands.

Third, many consumers cannot afford mega couture brands in the very beginning, so they rely on trendy styles from the low-end brands to build up popularity. Once the popularity is high enough, they start to be able to afford more high-end brands due to the cost decreasing effect of popularity. The three above-mentioned mechanisms are market expansion effect brought by fashion copycats. Our findings demonstrate that copycats' market expansion effect dominates the competition effect, leading to a positive net effect on the demand of high-end brands. In addition, the results also show a boost in the choice of compatibility, which for many bloggers is a substitute for distinctiveness.

3.7.1.1 Discussion on Firm Strategy

As we do not have firm side data and only model the consumers' decision making, our results speak to the cases where fast fashion firms do not find a way to dramatically change their styles and firms charge the original price. This section discusses how firms would react to the counterfactual policy and the corresponding demand effect on high-end brand.

Fast fashion firms would either put the same effort in designing their own styles, which is less distinctive and matches not well with higher-end brands, but decrease the price to attract more consumers. This escalated price competition would further decrease the demand for mega couture brands. Alternatively, the low-end brands may invest more in coming up a large number of their own styles to make them not only distinctive but also compatible with higher-end brands. We keep ignorant about how they can achieve this goal with the current low price, but one possibility is that they may transform into designer brands, providing better styles but also higher price due to the higher production cost. The demand effect of prohibiting copycats on mega couture brands is most likely still negative because there is only a small proportion of consumers mix and match designer brands with mega couture brands.

Regarding mega couture brands' strategic reaction, they may start a low-end product line, providing similar styles to those from high-end product line but charge lower price (i.e., umbrella branding). However, in this case, the low-end product line may erode the parent brand's value, and the demand effect is not clear without additional information and further empirical study. The other strategy the mega couture brands may use is to simply lower the price thus attract more demand. In this case, the mega couture brands' profit is definitely lower than before because they have to charge a lower price to achieve the same demand.

Overall, we can see if fast fashion copycats were prohibited, neither fast fashion firms nor mega couture brands can find an easy way to combat the loss.

3.7.1.2 Discussion on Generalizability

As we acknowledged upfront, an average consumer may differ from a fashion social media user. To generalize to all other consumers who do not use social media or do not post fashion content on social media, we need further study and additional data which is not available for now. However, we still can try to make the results more generalizable. In our sample, there are around 10% users are professional fashion bloggers who operate their own blogging website. They may have very different incentive than an average consumer. So we exclude these professional bloggers and check the choice change for other users. The results are reported in the second column of Table 13. Compared to the first column, we see the net effect is still negative but with smaller magnitude. This is due to the fact that professional bloggers care more about popularity, and the lack of fast fashion copycats makes it harder to build up popularity. As a result, the value increasing effect and cost decreasing effect of popularity cannot work their way to increase adoption of mega couture brands.

3.7.2 Offline Market: Absence of Follower Effect

This provides a similar scenario as the offline fashion market, where people do not have the "likes" information for what they wear. The posting decision is analogous to the purchase decision. There are two consequences. First, one's utility does not involve the likes' impact; second, the cost of purchasing will not decrease over time.

We seek to evaluate how less/more likely a blogger would post. This analysis allows us to understand how many more purchases can be achieved by the existence of social media, compared to the traditional offline market. Companies can leverage the "follower function" of social media to boost their revenue. The results in the fourth column of Table 13 show that there would be a 9.07% drop in post probability if bloggers did not know about others' following. High-end brands suffer most with the largest drop in choice probability, because the cost decreasing effect is also deprived, and this affects the purchase of high-end brands most.

3.7.3 When Brands Are Not Observable

As previously mentioned, our data context has a special brand tagging function, which requires users to pinpoint the brands for the clothing items in their fashion look and clearly lists the brands besides the picture. In most other social medium, there is no such feature. Moreover, in the offline market, brand logos are usually hidden or not obvious. In this counterfactual study, we investigate how fashion bloggers make choices if there were no way to inform others about the brand information. Consistent with the main analysis, when there is no brand information for a clothing item, the belief is that it is a high street brand (i.e., low-end brand). The results are shown in the third column of Table 13.

The results show that there would be 0.21% fewer posts. The users would post 0.74% more fast fashion brands, but 0.25% fewer designer brands and 0.30% fewer mega couture brands. Moreover, both style features are improved. Intuitively, when higher-end brands cannot be identified, bloggers cannot use high-end brands to attract peer likes anymore. Therefore, for those who care about popularity and peer likes a lot, they have fewer means to achieve the goal, resulting in lower utility from blogging and thus fewer posts. On the other hand, they switch focus to either using more low-end brands or investing more on styles. As the cost of a low-end brand is lower, bloggers would be able to purchase more low-end clothes, resulting in more posts of fast fashion brands. In summary, for higher-end brands, hiding or obscuring the brand logo would benefit the lower-end brands but hurt high-end brands, given all brands can provide similar styles. This result implies that, high-end brands may want to design their brand logos obvious enough to consumers, so that those who care about peer likes would have greater incentive to buy.

Table 13. Results of Counterfactual Studies

	Counterfactual Policies				
	No copycats	No copycats (Sub sample)	No brand info	No followers	Targeted ranking
Δ#Posts (%)	-7.92%	-6.12%	-0.21%	-9.07%	0.90%
Δ#Fast fashion (%)	-5.78%	-5.52%	0.74%	-6.07%	0.61%
Δ#Designer (%)	-7.68%	-6.08%	-0.25%	-11.78%	0.87%
Δ#Mega couture (%)	-12.44%	-9.32%	-0.30%	-19.86%	0.89%
ΔCompatibility (%)	-3.34%	-3.51%	99.93%	-	-
ΔDistinctiveness (%)	-23.87%	-15.76%	19.88%	-	-

3.7.4 Platform Ranking System: If "Sensitive" Users Are Prioritized

As bloggers value others' opinions differently, the website can change the ranking system so that it favors those "sensitive" bloggers, to incentivize more fashion posts. Effectively, the platform can exogenously change the number of likes a post can get through the ranking system. We run a counterfactual analysis on the top 10% of sensitive bloggers by exogenously giving one more like to their posts. The results in the last column of Table 13 show that, on average, there would be a 0.90% increase in the probability of posting in each period. There is a 0.61%, 0.87%, and 0.89% increase in the number of items posted from brand levels 1, 2, and 3 respectively.

3.8 Contribution and Limitations

In the fashion market, it has been argued that the fast fashion brands copy the designs of the high-end fashion brands. This practice can potentially reduce the distinctiveness of the luxury fashion brands thus erode their brand equity. However, there is no systematic study attempting to investigate the impact of fast fashion copycats on high-end brand equity and the underlying reasons. The key challenges limiting such study are the lack of scalable ways to quantify fashion styles and the unavailability of large-scale data on individuals' choices over brands and styles. In this paper, we use the user-generated data a large fashion sharing platform and the state-of-the-art deep learning methods on image analytics to quantify fashion styles. Given fashion social media users' significant impact on fashion trend and demand, understanding their decision process sheds light on fashion consumers' choices across population. For fashion goods, brand and style are two of the product attributes consumers care about most. Incentive-wise, fashion bloggers make the tradeoff between self-intrinsic tastes and building up popularity. To figure out the underlying reasons of how fast fashion may affect high-end brands' demand, we build a structural model to investigate fashion social media users' decision processes that reveals their heterogeneous responses to brands, styles, and popularity.

Our results show that styles and brands are quite substitutable for most people. In other words, they find it easy to substitute a high-end brand with good styles to derive the same utility. These are the consumers that high-end brands could lose to the low-end brands providing comparable styles. We also find that most users value being popular (or peer likes), but there exists a large variance in how much they value popularity. This variation explains why some people keep posting even though they get almost zero peer likes all the time. Moreover, we find that a higher popularity can help reduce posting cost which is consistent with the fact that fashion bloggers with lots of followers are sponsored by fashion companies to post about their products.

In the main counterfactual analysis, we restrict the availability of style choice for fast fashion brands. The results show that not only the fast fashion brands will suffer, but also the high-end brands will be worse off. This means that a more restrictive copyright law on fashion design may not necessarily help the mega couture brands. We found three mechanisms that contribute to this result. Because the more restricted choice of styles from low-end brands would limit the mix-and-match choices for consumers and put them on greater financial constraint to get a satisfactory ensemble of clothes, resulting in them buying less high-end brands. Moreover, the lack of good styles from low-end brands will make it harder for consumers to build up popularity/likeability, which limits the value of what they wear, including high-end brands. A low popularity also makes it unlikely to get a cost reduction from high-end brands. All these reasons indicate that copycats can benefit the high-end brand, as our findings suggest the market expansion effect dominates the competition effect. We also simulate the case where brand information cannot be seen, which is similar to the offline market where brand logos are hidden or other social media where brand tagging is not featured. The results indicate that on average, fast fashion benefits, whereas the high-end brands suffer. Another counterfactual analysis indicates that there would be a 9.07% drop in post probability if users did not know about peer likes, and high-end brands would suffer the most. This scenario is similar to that of the offline market. This analysis, therefore, allows us to understand how many more purchases can be achieved by the mere existence of social media, compared to the traditional offline market. We also find that an alternative ranking system that prioritizes people who more highly value "likes" can hugely increase the amount of content generated on the platform.

There are several limitations of the current paper that call for future studies. First, the style measure is an objective measure, whereas different consumers may evaluate the compatibility and distinctiveness differently. So the interpretation of the results should be based on the objective style measure we use. An ideal scenario is that we get to know how each consumer evaluates styles. Though this is not feasible, future research could do a more targeted analysis if there were more granular information about consumers from lab experiments or surveys and could then match people's evaluations of styles based on these observables. Second, though we tried to capture the most salient factors in a fashion post—that is, the model face, the body feature, and the style of clothing items—we did not include everything that may affect the likes a post can get. For example, the accessories may matter. The challenge is that although the detection algorithms for fashion items are state-of-the-art, they are not perfect, especially for items that are too small (e.g., earrings and hats are comparatively quite small). Had we been able to use a better detection algorithm or rich training set, we could have incorporated more factors in the model.

Reference

Aguirregabiria, B Y Victor, and Pedro Mira. 2007. "Sequential Estimation of Dynamic Discrete Games." *Econometrica* 75 (1): 1–53.

Amaldoss, Wilfred, and Sanjay Jain. 2005. "Conspicuous Consumption and Sophisticated Thinking." *Management Science* 51 (10): 1449–66.

Amaldoss, Wilfred, and Sanjay Jain. 2005. "Pricing of Conspicuous Goods: A Competitive Analysis of Social Effects." *Journal of Marketing Research* 42 (1): 30–42.

Amos, Brandon, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. 2016. "OpenFace: A General-Purpose Face Recognition Library with Mobile Applications." *Technical Report CMU-CS, CMU School of Computer Science* 16 (118).

Arcidiacono, Peter, and Robert Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity." *Econometrica* 79 (6): 1823–67.

Bernhein, Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5): 841–77.

Borkovsky, Ron N., Avi Goldfarb, Avery M. Haviv, and Sridhar Moorthy. 2017. "Measuring and Understanding Brand Value in a Dynamic Model of Brand Management." *Marketing Science* 36 (4): 471–99.

Brewer, Marilynn. 1991. "The Social Self: On Being the Same and Different at the Same Time." *Personality and Social Psychology Bulletin* 17 (5): 475–82.

Fenigstein, A, M. F Scheier, and A.H Buss. 1975. "Public and Private Self Consciousness: Assessment and Theory." *Journal of Consulting and Clinical Psychology* 43 (4): 522–27.

Girshick, Ross. 2015. "Fast R-CNN." *arXiv:1504.08083v2 [cs.CV]*.

Goldfarb, Avi, Qiang Lu, and Sridhar Moorthy. 2009. "Measuring Brand Value in an Equilibrium Framework." *Marketing Science* 28 (1): 69–86.

Ha, Yu-I, Sejeong Kwon, Meeyoung Cha, and Jungseock Joo. 2017. "Fashion Conversation Data on Instagram." *International AAAI Conference on Web and Social Media. arXiv:1704.04137v1 [stat.ML]* 13 Apr.

Hadsell, Raia, Sumit Chopra, and Yann LeCun. 2006. "Dimensionality Reduction by Learning an Invariant Mapping." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2: 1735–42.

Hartmann, Wesley R., and Harikesh S. Nair. 2010. "Retail Competition and the Dynamics of Demand for Tied Goods." *Marketing Science* 29 (2): 366–86.

Helmore, Edward. 2010. "Google Targets Fashion Market." *Guardian*.

Horen, Femke van, and Rik Pieters. 2012. "When High-Similarity Copycats Lose and Moderate-Similarity Copycats Gain: The Impact of Comparative Evaluation." *Journal of Marketing Research* XLIX (February): 83–91.

Hotz, V Joseph, and Robert a Miller. 1993. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *The Review of Economic Studies* 60 (3): 497–529.

Hu, Y., L. Manikonda, and S. Kambhampati. 2014. "What We Instagram : A First Analysis of Instagram Photo Content and User Types." *Proceedings of the Eight International AAAI Conference on Weblogs and Social Media*.

Imai, Susumu, Neelam Jain, and Andrew Ching. 2009. "Bayesian Estimation of Dynamic Discrete Choice Models." *Econometrica* 77 (6): 1865–99.

Kamakura, Wagner A, and Gary J Russell. 1993. "Measuring Brand Equity with Scanner Data." *International Journal of Research in Marketing* 10 (1): 9–22.

Keller, Kevin Lane, and Donald R. Lehmann. 2006. "Brands and Branding: Research Findings and Future Priorities." *Marketing Science* 25 (6): 740–59.

Kocabey, Enes, Mustafa Camurcu, Ferda Ofli, Yusuf Aytar, Javier Marin, Antonio Torralba, and Ingmar Weber. 2017. "Face-to-BMI: Using Computer Vision to Infer Body Mass Index on Social Media." *arXiv:1703.03156v1 [cs.HC]*.

Kuksov, Dmitri. 2007. "Brand Value in Social Interaction." *Management Science* 53 (10): 1634–44.

Liu, Liu, Daria Dzyahura, and Natalie Mizik. 2018. "Visual Listening In: Extracting Brand Image Portrayed on Social Media." *Working Paper*.

Liu, Ziwei, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. 2016. "DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, no. 1: 1096–1104.

Ma, Liye, Alan L Montgomery, and Michael D Smith. 2016. "The Dual Impact of Movie Piracy on Box - Office Revenue : Cannibalization and Promotion." *Working Paper*.

Nam, Hyoryung, Yogesh V. Joshi, and P.K. Kannan. 2017. "Harvesting Brand Information from Social Tags." *Journal of Marketing* 81 (4): 88–108.

Oberholzer-Gee, Felix, and Koleman Strumpf. 2007. "The Effect of File Sharing on Record Sales: An Empirical Analysis." *Journal of Political Economy* 115 (1): 1–42.

Qian, Yi. 2008. "Impacts of Entry by Counterfeiters." *The Quarterly Journal of Economics* 123 (4): 1577–1609.

Qian, Yi. 2014. "Counterfeiters: Foes or Friends? How Counterfeits Affect Sales by Product Quality Tier." *Management Science* 60 (10): 2381–2400.

Simmel, Georg. 1904. “Fashion.” *International Quaterly* 10 (22): 130–55.

Smith, M D, and R Telang. 2009. “Competing with Free: The Impact of Movie Broadcasts on DVD Sales and Internet Piracy.” *MIS Quarterly* 33 (2): 321–38.

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, et al. 2015. “Going Deeper with Convolutions.” *CVPR*.

Toubia, Olivier, and Oded Netzer. 2017. “Idea Generation, Creativity, and Prototypicality.” *Marketing Science* 36 (1): 1–20.

Veit, Andreas, Balazs Kovacs, Sean Bell, Julian Mcauley, Kavita Bala, and Serge Belongie. 2015. “Learning Visual Clothing Style with Heterogeneous Dyadic Co-Occurrences.” *arXiv:1509.07473v1 [cs.CV]*.

Wang, Quan, Beibei Li, and Param Vir Singh. 2018. “Copycats versus Original Mobile Apps: A Machine Learning Copycat Detection Method and Empirical Analysis.” *Information Systems Research* 29 (2): 273–91.

Yoganarasimhan, Hema. 2012. “Cloak or Flaunt? The Fashion Dilemma.” *Marketing Science* 31 (1): 74–95.

Appendix

3.9 More on summary statistics

Table 14 shows how many times each brand appears in the posts in total.

Table 14. Frequency of brands (Top 60)

Brand	Count	Brand	Count	Brand	Count	Brand	Count
H&M	35891	Dr. Martens	2743	GUESS	1490	Celine	895
Zara	28865	Topman	2546	Adidas	1393	Prada	876
Forever 21	15384	Stradivarius	2376	Diy	1382	Choies	847
Topshop	10273	River Island	2325	Cheap Monday	1359	Blanco	793
Vintage	8493	Aldo	2265	Shein	1306	Gucci	788
Mango	5518	Pull&Bear	2169	Uniqlo	1288	J. Crew	783
Primark	5504	Nike	2160	Marc by Marc Jacobs	1258	Ralph Lauren	776
Asos	5329	Thrifted	2051	Louis Vuitton	1156	C&A	772
American Apparel	5006	Jeffrey Campbell	2040	Cotton On	1139	Vero Moda	768
Converse	3922	Vans	1978	New Yorker	1106	Old Navy	755
Ray-Ban	3716	Chanel	1836	Gina Tricot	1053	Nine West	749
Levi’s(r)	3597	Gap	1644	Monki	957	Calvin Klein	742
Urban Outfitters	3450	Steve Madden	1575	OASAP	908	Pimkie	732
New Look	3271	Romwe	1564	Target	902	Alexander Wang	732
Bershka	3169	Michael Kors	1523	FrontRowShop	901	Saint Laurent	671

3.10 B. Survey

We are interested in assessing the aesthetic quality of human photos on three factors: clothing compatibility, facial attractiveness and body attractiveness. These are subjective measures, but you can guide your judgments according to the following instruction and examples.

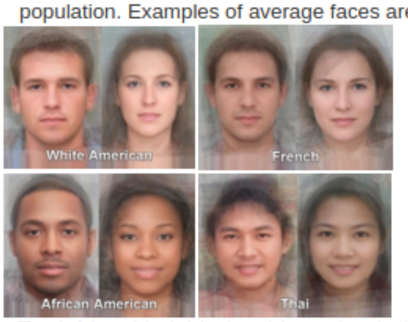
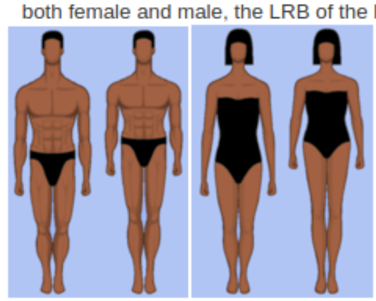
Instruction.

1. Clothing Compatibility: how well the clothing items match together.

Example.

The 1st item is compatible with the 2nd item but not compatible with the 3rd item.

2. Facial (Body) attractiveness is the degree to which a person's facial (body) features are considered aesthetically pleasing or beautiful. Below are some general aspects you can refer to.

Facial attractiveness	Body attractiveness
· Symmetry. Symmetric images are usually viewed as more attractive than asymmetric images. For example, (a) is more attractive than (b) in the following picture.  · Beauty of the average. A face is usually considered more attractive if it more closely resembles the majority of other faces within a population; non-average faces have more extreme characteristics than the average of a population. Examples of average faces are shown below:  · Facial health (skin health and color). High healthiness is typically associated with higher score of attractiveness. As shown in the example below, the right face is more attractive than the left face. 	· Symmetry. Similar to a face, a symmetric body is viewed as more attractive than an asymmetric one. · Body Health. Healthy bodies are considered more attractive. For example, overweight or emaciated bodies are considered of low attractiveness. All the following aspects also affect a body's attractiveness. However, preference varies across different cultures and regions. · Leg-to-body ratio (LBR). In the following pictures for both female and male, the LRB of the left body is lower.  · Waist-to-hip ratio. · Breast-to-waist ratio. · Height.

Examples.

Below we offer some examples of faces and bodies with different attractiveness levels (1 to 10). Please look at these examples to set your expectations.

Examples of faces with various attractiveness levels in randomized order (1st row for women, 2nd row for men)

Examples of bodies with various attractiveness levels in randomized order (1st row for women, 2nd row for men)

Questions.

[Picture given here]

Please answer the following questions regarding the given picture. For the first two questions, please disregard clothing styles and focus only on face (body) attractiveness:

[Compatibility] Do you think the clothing items in the photo are compatible (match well)?
Yes/No

[Body attractiveness] How attractive do you think the person's **body** is? (1: least attractive, 10: most attractive)
(Scale 1~10)

[Facial attractiveness] How attractive do you think the person's **face** is? (1: least attractive, 10: most attractive)
(Scale 1~10)

[Picture aesthetics] How visually pleasing do you think the whole picture is? (1: least visually pleasing, 10: most visually pleasing)
(Scale 1~10)

[Clothing Fashion] How fashionable is the clothing style? (1: worst fashion, 10: best fashion)
(Scale 1 ~10)

[Clothing Price Appearance] How expensive do you think the clothings are? (1: very cheap, 10: very expensive)
(Scale 1 ~10)

[Subject Gender] Is the person in the picture a female or male?
Male/Female

[Subject Glasses] Is the person wearing glasses in this picture?
Yes/No

What is **your** age?
A: <20
B: 20~29
C: 30~39
D: 40~49
E: 50~60
F: >60

What is **your** gender?
Female/Male

What is **your** ethnicity?
A) White
B) Black or African American
C) East Asian (e.g., Chinese, Korean, Japanese)
D) Other Asian
E) Hispanic or Latino
F) Other

3.11 Solve for optimal style choices

For a given targeted $\widehat{Like}_t$, which affects the future utility, each individual would choose styles that maximize their current utility. That is, user i solves the following optimization problem (ignoring the subscript i):

$$\max_{(f_1, f_2)} \; [\beta {f_1}^{\rho_2} + \; (1 \; - \; \beta){f_2}^{\rho_2}]^{\frac{1}{\rho_2}}$$

s.t. $b_1 f_1 + \; b_2 f_2 = \mathrm{L}$

where L is a constant given by

$$L = \widehat{Like}_t - \vec{\mathbf{b}} \cdot (x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, Like_{t-1}, face, body, age, gender),$$

and $\vec{\mathbf{b}}$ is the coefficients estimated from the state transition regression $\widehat{g}\,(\cdot)$.

For the Lagrangian

$$H = [\beta {f_1}^{\rho_2} + \; (1 \; - \; \beta){f_2}^{\rho_2}]^{\frac{1}{\rho_2}} \; + \; \lambda[L \; - \; (b_1 f_1 + \; b_2 f_2)]\;)$$

The first order conditions are

$$L_{f_1} = \frac{1}{\rho_2}[\beta {f_1}^{\rho_2} + \; (1 \; - \; \beta){f_2}^{\rho_2}]^{\frac{1}{\rho_2}} \beta \rho_2 {f_1}^{\rho_2 - 1} - \lambda b_1 = 0$$

$$L_{f_2} = \frac{1}{\rho_2}[\beta {f_1}^{\rho_2} + \; (1 \; - \; \beta){f_2}^{\rho_2}]^{\frac{1}{\rho_2}} (1-\beta) \rho_2 {f_2}^{\rho_2 - 1} - \lambda b_2 = 0$$

Solving the above system of equations, we have

$$\frac{b_1}{b_2} = \frac{\beta {f_1}^{\rho_2 - 1}}{(1-\beta){f_2}^{\rho_2 - 1}}$$

$\Longrightarrow$

$$f_2 = f_1 \left[\frac{b_2 \beta}{b_1(1-\beta)}\right]^{\frac{1}{\rho_2 - 1}}$$

Plug in $L_\lambda = 0$, we have

$$f_1^* = \frac{L}{b_1 + b_2 A}, \qquad f_2^* = \frac{L \cdot A}{b_1 + b_2 A}$$

where $A = \left[\frac{b_2 \beta}{b_1(1-\beta)}\right]^{\frac{1}{\rho_2 - 1)}}$.

3.12 Estimation Algorithm.

Let *I* denote the total number of bloggers, *N* is the number of previous iterations used for calculating the expected value for the current iteration.

1. At iteration r, the state-depend value $\tilde{V}$and heterogenous parameters $\boldsymbol{\Theta_i}$ in the past N iterations are $H^r = \{\{\boldsymbol{\Theta_i^m}, \tilde{V}^m(S_i^m; \boldsymbol{\Theta_i^m})\}_{i=1}^{I}\}_{m=r-N}^{r-1}$, where $\theta_i \in \boldsymbol{\Theta_i} = \{\alpha_{i1}, \alpha_{i2}, \boldsymbol{\rho_i}, \boldsymbol{\eta_i}, \boldsymbol{t_i}, \boldsymbol{\gamma_i}, \delta_i\}$.

2. Draw μ_θ^r, the population mean of θ_i, from the posterior distribution based on σ_θ^{r-1} and parameters estimated in the last iteration $\{\theta_i^{r-1}\}_{i=1}^{I}$, i.e., $\mu_\theta^r \sim N\left(\frac{\sum_{i=1}^{I}\theta_i^{r-1}}{I}, \sigma_\theta^{r-1}\right)$.

3. Draw σ_θ^r, the population variance of θ_i, from the posterior distribution based on the updated μ_θ^r and $\{\theta_i^{r-1}\}_{i=1}^{I}$, i.e., $\sigma_\theta^r \sim IG\left(\frac{I}{2}, \frac{\sum_{i=1}^{I}(\theta_i^{r-1}-\mu_\theta^r)^2}{2}\right)$.

4. Draw new parameters θ_i^r for each individual $i = 1, \dots, I$ from the posterior distribution $f_i(\theta_i^r | \mu_\theta^r, \sigma_\theta^r, St_i{}^o, Br_i{}^o, P_i{}^o) \propto \pi(\theta_i | \mu_\theta^r, \sigma_\theta^r) L\{St_i{}^o, Br_i{}^o, P_i{}^o | \theta_i^r\}$, where $St_i{}^o, Br_i{}^o, P_i{}^o$ are the observed choices of style, brand, and post.

We use Metropolis-Hastings algorithm to draw from the above posterior distribution.

(1). Draw candidate parameters θ_i^{*r} from a proposal density $q(\theta_i^{r-1}, \theta_i^{*r})$, essentially adding some perturbation to θ_i^{r-1}, for example, $\theta_i^{*r} \sim N(\theta_i^{r-1}, \varepsilon^2)$.

(2). Given θ_i^{*r}, compute the likelihood, i.e., $L\{St_i{}^o, Br_i{}^o, P_i{}^o | \theta_i^{*r}\}$. Computation of the likelihood with the observed continuous choice ($St_i{}^o$) is traditionally done in two ways: first, discretizing the continuous choice space into countable discrete choices; second, numerical approximation using kernel smoothing (e.g., Yao et al., 2012; Liu et al., 2018). The downside of the first way is the loss of the continuous nature of the corresponding choice. The second approach reserves continuity but requires thousands of draws of random errors and solving for optimal choices, which could be computationally very costly, especially for the case without closed-form solutions.

Our approach borrows the spirits of both ways. On the one hand, we reserve the continuous nature of the style choices $St_i{}^o$, that is, we allow the choices to take any positive real numbers. On the other hand, to alleviate the computational burden, we also use some numerical approximation or smoothing, based on the 'discretizing' the number of likes which naturally take discrete values (i.e., integers).

Specifically, for a given target likes $\widehat{Like}$, there exists a unique choice $St^*{}_i$ (please see Appendix C for the optimal solution). The $\widehat{Like}$ can take any integer from 0 to the maximum of likes achieved across all the posted fashion looks $K = 1747$. The probability for the observed style choices $St_i{}^o$ that lead to $\widehat{Like} = \hat{g}(St_i{}^o, Br_i{}^o, S_{it})$ is therefore

$$Pr\{St_i{}^o, Br_i{}^o, P_i{}^o | S_{it}, \theta_i^{*r}\} = \frac{\exp\left\{v^r\left(\widehat{Like}, Br_i{}^o, P_i{}^o | S_{it}, \theta_i^{*r}\right)\right\}}{\exp\{v^r(P_i{}^o = 0 | S_{it}, \theta_i^{*r})\} + \sum_{k=0}^{K}\sum_{n=1}^{35}\exp\left\{v^r\left(\mathrm{k}, Br_{i,n}, P_i{}^o = 1 | S_{it}, \theta_i^{*r}\right)\right\}}$$

where $v^r\left(\widehat{Like}, Br_i{}^o, P_i{}^o | S_{it}\right) = v^r\left(St_i{}^o, Br_i{}^o, P_i{}^o | S_{it}\right)$ if $\widehat{Like} = \hat{g}(St_i{}^o, Br_i{}^o, S_{it})$ is an integer. Otherwise, $v^r\left(\widehat{Like}, Br_i{}^o, P_i{}^o | S_{it}\right) = v^r\left(\widehat{Like\_nbr}, Br_i{}^o, P_i{}^o | S_{it}\right) = v^r\left(St^*{}_i, Br_i{}^o, P_i{}^o | S_{it}\right)$, where

$\widehat{Like\_nbr}$ is the nearest integer neighbor of $\widehat{Like}$, and $St^*{}_i$ is the optimal choices to achieve the target $\widehat{Like\_nbr}$.

The choice specific value function $v^r(St_i{}^o, Br_i{}^o, P_i{}^o | S_{it})$ is the per-period utility plus the expected future value $EV^r(S_i)$, calculated with a weighted average of the past state-specific value $\{\{\tilde{V}^m(S_i^m; \boldsymbol{\Theta_i^m})\}_{i=1}^I\}_{m=r-N}^{r-1}$. The cumulative number of likes, denoted by s, as the stochastically evolving state, follow the state transition probability $T(St_i, Br_i, P_i | s, \hat{\sigma})$. We have the estimated standard deviation for the random error $\zeta_{it} \sim N\left(0, \hat{\sigma}^2\right)$, resulting from the transition regression.[50] With a step size of 1, the probability of getting y likes rather than $\widehat{Like}$ is therefore $1 \times \phi_{(0,\hat{\sigma})}(y - \widehat{Like})$, where $\phi_{(0,\hat{\sigma})}(\cdot)$ is the density function for normal distribution $N\left(0, \hat{\sigma}^2\right)$. So, we have

$$EV^r(St_i, Br_i, P_i, S_i) = \sum_{m=r-N}^{r-1} \tilde{V}^m(S_i^m; \boldsymbol{\Theta_i^m}) \frac{K_{h\theta}(\theta_i^{*r} - \theta_i^m) K_{hS}(s_i^{m\prime} - s_i)}{\sum_{j=r-N}^{r-1} K_{h\theta}\left(\theta_i^{*r} - \theta_i^j\right) K_{hS}\left(s_i^{j\prime} - s_i\right)}.$$

Then we can calculate the likelihood $L\{St_i{}^o, Br_i{}^o, P_i{}^o | \theta_i^{*r}\}$.

(3) Repeat the above to obtain the likelihood for the old parameter from the last iteration $L\{St_i{}^o, Br_i{}^o, P_i{}^o \middle| \theta_i^{r-1}\}$.

(4) Having obtained the likelihood, we can determine whether to accept the candidate parameters $\hat{\theta}_i^{*r}$, with acceptance probability λ given by

$$\lambda = \min\left\{\frac{\pi(\theta_i^{*r} | \mu_\theta^r, \sigma_\theta^r) L\{St_i{}^o, Br_i{}^o, P_i{}^o | \theta_i^{*r}\} q\left(\theta_i^{*r}, \hat{\theta}_i^{r-1}\right)}{\pi\left(\theta_i^{r-1} | \mu_\theta^r, \sigma_\theta^r\right) L\{St_i{}^o, Br_i{}^o, P_i{}^o \middle| \theta_i^{r-1}\} q(\theta_i^{r-1}, \theta_i^{*r})}, 1\right\}.$$

Repeat (1) ~ (4) for all individuals $i = 1, \ldots, I$.

5. Given the accepted parameters resulting from step 4, θ_i^r, we can update the value function for each individual, $\{\tilde{V}^r(S_i^r; \boldsymbol{\Theta_i^r})\}_{i=1}^I$.

Given the Type-1 extreme value distribution for the brand-and-style specific random error, the value takes the following form

$$\tilde{V}^r(S_i^r; \boldsymbol{\Theta_i^r}) = 0.577 + \log\left[\exp\{v^r(P_i{}^o = 0 | S_i, \theta_i^r)\} + \sum_{k=0}^{K} \sum_{n=1}^{35} \exp\left\{v^r\left(\mathrm{k}, Br_{i,n}, P_i{}^o = 1 | S_i, \theta_i^r\right)\right\}\right]$$

6. Proceed to the next iteration $r + 1$ and repeat the above steps until convergence.

[50] An ordinal logistic regression was also tested for robustness check.