Selling a Viral Product

Yating Yuan*

July 17, 2024

Abstract: Sellers often launch a new product with different price-quality packages (versions) when social learning is prevalent. Yet few papers explain the seller's versioning incentive from a social learning perspective. This paper explores when and how observational learning incentivizes a monopolist to sell different versions even in a common value setting. The dynamic learning process sheds new light on versioning beyond traditional explanations. The findings highlight the pivotal role of consumers' private information quality in determining two distinct selling strategies. In markets with noisy private signals, the seller offers a single version; with precise signals, she adds a cheaper basic version.

Keywords: versioning, observational learning, nonlinear pricing

JEL classification: D42 D83 L11 L15

---

*Department of Economics, University of Warwick, Coventry, CV47AL, United Kingdom. Email: yating.yuan@warwick.ac.uk. I am grateful to my supervisors Motty Perry and Jacob Glazer for their guidance, support, and insightful comments. I would also like to thank Guo Bai, Ilan Kremer, Phil Reny, Costas Cavounidis, Daniele Condorelli, Ravi Jagadeesan, Raghav Malhotra, Kirill Pogorelskiy, Francesco Squintani, Ao Wang, Yu Fu Wong and participants at microeconomics work-in-progress workshops at Warwick, the 2023 RES PhD Conference, the 2023 Lisbon Meeting of Game Theory, the 2023 HEC Paris PhD conference, the 2023 AMES at Singapore, the 2023 Asian School of Economic Theory, and the 2023 EARIE conference, NASMES 2024, and EWET 2024 for helpful comments and suggestions.

1 Introduction

Do firms profit from selling multiple versions when launching a new product? The economic literature has provided various explanations. However, most existing theories are static in nature, while in a digital era, new product markets often feature a dynamic observational learning process. When Apple launched the first iPhone in 2007, no one knew exactly how much value it would bring to everyday life. Over time, the public came to understand the value through the growing number of users. It is such public belief, rather than private tastes, that drives market demand in the long run. What is the optimal menu for the firm to launch such a novel product? When is it profitable to sell multiple versions, like iPhone and iPhone Pro, instead of a single version?

To address these questions, this paper proposes a dynamic model in which consumers share a common value but arrive sequentially with different information, and offers a new justification for the multi-version policy. In the presence of observational learning, a cheap basic version serves as a tool to guide social learning in the short run. It allows the first batch of consumers to try at a cheap price. A monopolist then gets to sell a much more expensive premium version, which would not have attracted any purchasers if it were the only option on the market. However, this role of versioning does not always work. When the surplus from learning is minimal and consumers have noisy private information, a single-version menu is optimal. Otherwise, it is profitable to add a basic version.

Reward-based crowdfunding exemplifies such new product markets. It is common practice for sellers to launch their crowdfunding campaigns with different price-quality packages[1] and consumers observe the number of pur-

[1]For instance, a comic book writer may offer both a digital version and a paperback version of their new book at different prices. Similarly, a game designer might sell a wide range of packages. Expensive ones often offer premium features to enhance the gaming experience.

chases (backers) for each package. The multi-version policy is also widely adopted by firms selling online service products, such as Grammarly, ChatGPT, and D-ID Studio. Their motivation is precisely to let the cheap version go viral online. Once the market recognizes its core value, it will hopefully be ready for a premium version. With a simple and tractable model, this paper examines a monopoly pricing and versioning problem from the new Internet era, providing fresh insights into optimal selling strategies in various online markets.

A monopolist releases a new product for which consumers share an unknown binary common value. The seller aims to maximize expected long-run average profits. She can offer a single-version or two-version menu with exogenous, observable vertical qualities and (fixed) prices[2]. An infinite number of consumers then arrive one at a time, each with a private signal about the value. They also observe their predecessors' decisions and obtain public information from there. At the end of each period, they either choose a version to buy or abstain. Consumers become increasingly informed over time until an informational cascade occurs.[3]

The article first examines the optimal price in a single-product benchmark model, focusing on the relationship between the optimal price, long-run demand elasticity and private signal informativeness. It then discusses the optimal learning schemes and when to offer two versions.

In the single-product benchmark, charging a higher price increases the margin (price effect) but reduces the probability of a buy cascade in the limit (quantity effect). The relative magnitude of these effects, and hence

[2]The fixed pricing assumption applies to many real-life situations. Many firms do not change prices frequently due to managerial inattention and brand image concerns (Arcidiacono et al. (2020), DellaVigna and Gentzkow (2019), Reimers and Waldfogel (2017), Phillips (2015)). Coca-Cola faced a huge backlash from consumers in 1999 when their then CEO Ivester considered introducing a new vending machine that changed price with the outside temperature. Most crowdfunding projects offer a static price for each package. Grammarly's pricing plan has remained unchanged for years.

[3]The observational learning part of the model is built on Smith and Sørensen (2000).

the price elasticity of demand, is determined by the informativeness of the private signal. With a noisy signal, the quantity effect dominates and a low price is optimal. To stay safe, the seller may even set a low price that triggers an immediate 'buy cascade'[4]. A precise signal, however, weakens the quantity effect, incentivizing the seller to choose a high price. In doing so, she bets on the prior probability that the core product is good and reaps the fruit of learning as consumers become increasingly optimistic.

The two-version model introduces another layer of trade-off. Now the question boils down to whether to introduce a cheaper basic version. This is because, under the assumption of zero production cost, the seller will always keep the premium product on the market. Imposing a higher production cost on the premium product only strengthens the seller's incentive to offer a basic version. The main results remain unchanged qualitatively.

As usual, adding a cheaper basic version brings information rent. The seller must lower the premium version's price to prevent high-belief consumers from purchasing the basic version. What is novel is on the benefit side.

Imagine consumers receive extremely precise signals. As discussed in the benchmark case, the seller's top priority now is to charge a high price as long as consumers do not immediately enter a no-buy cascade. Introducing a basic version is profitable here because it allows the first several consumers to learn at a relatively low price. This expands the learning set and eases the pressure of starting the process too close to a no-buy cascade, allowing for a further increase in the premium version's price.

A noisy signal, however, encourages the seller to stay safe by charging a low price on the premium version. Introducing an even cheaper basic version to the market only introduces unnecessary information rents. As such,

[4]Technically speaking, the buy cascade starting from the beginning is not an informational cascade but just a sequence of buy decisions.

a single-version menu is optimal. Theorem 2 provides a threshold characterization of how signal informativeness determines the seller's decision to expand the menu.[5]

The theoretical framework provides a foundation for future empirical analysis of the interaction between observational learning and the optimal menu. It highlights a novel relationship between private information quality and selling strategy, which has been less studied in the literature. The next section provides a more detailed discussion.

2 Related literature

An important early paper that examines the monopoly (fixed) pricing problem with observational learning is Welch (1992). In contrast to my model, Welch assumes a uniformly distributed state and a finite number of agents. Moreover, his signal structure is relatively noisy. The posterior expected value updates so slowly that the issuer (seller) will underprice to completely avoid a rejection cascade. My paper adds to the literature by demonstrating that when the private signal becomes precise, it is optimal to charge a high price and risk a rejection cascade. In such cases, a multi-version menu yields even higher profits.

Bose, Orosel, Ottaviani, and Vesterlund (2006, 2008) investigate a dynamic pricing problem in the informational cascade setting with finite

[5] As the main results highlight the role of private signal informativeness, it would be helpful to give some real-world examples where consumers may receive private information of varying precision. The consumer's private information is generally noisier if the core product is innovative. In the crowdfunding context, a consumer's private evaluation of the first book of a new series will not be as precise as their evaluation of the 10th book in the series. Another good example is the market for new treatments, as discussed in (Arieli, Koren, and Smorodinsky, 2022). A doctor gathers information by using limited free samples from pharmaceutical companies within the doctor's patient community. The realized success rate then gives each doctor a private signal about the value of the new treatment. The signal tends to be more precise if doctors receive more free samples, the samples are similar in effectiveness to the majority of products, or their patient communities are more diversified.

signals (binary in Bose et al. (2008)). Their setup bears similarities to my multi-version model. Because offering multiple versions with different prices at the outset seems to be a static alternative to dynamically adjusting the price of a single version. However, dynamic pricing offers the seller more flexibility to adjust their strategy as the learning process progresses. This paper offers new insights into the optimal menu design in this more constrained environment.

Several other studies have explored the optimal selling strategy in the presence of social experimentation. For instance, Bonatti (2011) investigates optimal dynamic menus for selling new experience goods to consumers with both a common value component and private taste. Laiho and Salmi (2021) considers a dynamic pricing problem when consumers can delay purchases. Bergemann and Välimäki (2002)[6] explores how dynamic competition among firms affects pricing and entry decisions. This literature assumes sales generate information via experimentation and simplifies the learning process to a publicly observable Brownian motion process. The assumption of public signal makes it difficult to discuss how private signal quality shapes the optimal selling strategy.

Another closely related field of study is menu pricing, which traces back to seminal works by Stigler (1963), Adams and Yellen (1976), Mussa and Rosen (1978) and Stokey (1979). Since then, numerous articles have explored conditions under which a monopolist prefers a multiple-item menu over a single-item menu (Salant, 1989; Anderson and Dana, 2009). A recent paper by Sandmann (2023) highlights the role of consumers' risk preferences. The evolving bundling literature (Haghpanah and Hartline, 2021; Ghili, 2023; Yang, 2024) establishes conditions for the optimality of pure and nested bundling. The conditions rely on relative values for and

[6] Other related papers include Bergemann and Välimäki (1997), Bergemann and Välimäki (2000), and Bergemann and Välimäki (2006)

sold-alone quantities of different goods. This screening literature typically assumes consumers' willingness to pay is exogenous. In contrast, my paper features endogenously formed variations in consumers' preferences due to observational learning. It allows me to unravel a novel observation on versioning and private signal quality.

Many empirical papers have already explored observational learning in various contexts, such as kidney exchange markets (Zhang, 2010), microloan markets (Zhang and Liu, 2012), music platforms (Newberry, 2016), and housing markets (Fan, Weng, Zhou, and Zhou, 2023). However, few have examined the effect of consumers' private information quality on the seller's pricing and versioning choices. According to Zhang and Liu (2012), investors in microloan markets behave differently during the learning process when they find their predecessor's private information is more precise. Based on this observation, the theoretical model developed here could guide further study of optimal selling strategies in such markets, especially when private information quality varies across different products.

3 Model: Monopoly Problem with Observational Learning

A monopolist seller plans to launch a new product. An infinite number of short-lived agents with unit demand arrive one at a time. $t \in \{1, 2, ..., \infty\}$.

State of the world. The product's core content has a binary value $V_\omega \in \mathbb{R}_+$ where $\omega \in \Omega := \{0, 1\}$. We assume $V_1 > V_0 \geq 0$. Neither the seller nor the agents observe the realized value. They share a common prior $\mu_1 := \Pr(\omega = 1) \in (0, 1)$.

Actions. The seller offers a menu with up to two options: a basic version $L = (p_L, q_L)$ with version quality $q_L \in \mathbb{R}_{++}$, and a premium version

$H = (p_H, q_H)$ with $q_H \in (q_L, \infty)$. These version qualities are vertical, exogenously fixed and perfectly observable.[7] The seller chooses a price schedule $\boldsymbol{p} := (p_L, p_H) \in \mathbb{R}^2_+$ which will be fixed over time. In each period t, an agent either buys a version or abstains. $\forall t, a_t \in \{L, H, r\}$.

Payoff. Agents receive a quasi-linear payoff from purchasing a version of the new product: $U(a_t, p, \omega) = u(q_i, V_\omega) - p_i$, for $a_t = i \in \{L, H\}$. The function $u : \mathbb{R}^2 \to \mathbb{R}$ is twice continuously differentiable, strictly increasing, and strictly supermodular in (q_i, V_ω). Let $u_{i\omega} := u(q_i, V_\omega)$ and $\bar{u}_i := u_{i1} - u_{i0}$ denotes the utility difference across states for each version i. The outside option $(a_t = r)$ gives zero payoff.

The supermodularity assumption implies that the marginal utility from observable version quality q_i increases with the core value V_ω. It is consistent with many real-life scenarios. For example, consumers are willing to pay more for physical comic book copies if they appreciate its core contents. iPhone users would prefer a longer battery life only when they enjoy the core design of the iPhone. Gamers become more interested in additional game features when they appreciate its graphics and gameplay design.

The risk-neutral seller maximizes her long-run average profits

$$\pi(a, \boldsymbol{p}) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} [\mathbf{1}(a_t = L)p_L + \mathbf{1}(a_t = H)p_H]$$

where $a := (a_1, a_2, ...)$ denotes the agents' action profile.

Signal structure. Each agent, upon arrival, receives a private signal s_t. It is i.i.d. across t and follows a differentiable distribution G^ω. Following Smith and Sørensen (2000), we map each signal realization into a 'private (posterior) belief' given a flat prior, $\tau(s) := \frac{g^1(s)}{g^1(s)+g^0(s)} \in (0, 1)$, where g^ω

[7] The observable qualities q_H and q_L differ from the unobservable value of the core product V_ω. Consider a comic book; both digital and physical copies share the same core content. While the quality difference between versions is obvious, whether the core content matches public taste is uncertain.

denotes the density functions of s_t. The private belief process $\langle \tau_t \rangle$ is then i.i.d. with a differentiable conditional cumulative distribution function F^ω. Let f^ω denote the associated density function. We assume no private signal fully reveals the state, which ensures F^0 and F^1 are mutually absolutely continuous with a common support $supp(F)$. Also, the private belief is bounded, i.e., $co(supp(F)) := [\underline{b}, \bar{b}] \subset [0, 1]$.

Timing.

$t = 0$: Nature chooses the state ω, and the seller sets a price schedule $p = (p_L, p_H)$ without observing the realized state.

$t = 1, 2, ..., \infty$: An agent arrives. She observes the price schedule p, decisions of previous agents, i.e., the public action history $h_t := (a_1, a_2, ..., a_{t-1})$, and a private signal s_t. She then chooses a version from the menu or walks away with nothing.

Equilibrium concept. The analysis focuses on pure-strategy perfect Bayesian equilibrium. In cases where an optimal menu does not exist, we will find a sequence of ε-optimal menus that converge to a well-defined limit menu. The comparative statics results are built on the associated limit profit maximization problem.

Definition (ε-optimal menu) For any real non-negative number ε, an ε-optimal menu is a price schedule $\boldsymbol{p}^\varepsilon$ that the seller cannot obtain more than ε in expected payoff by deviating from it.

Discussion. With **unbounded private beliefs**, the interesting interplay between pricing and learning diminishes because the long-run demand[8] will always equal the prior μ_1. Thus, from a long-run perspective, we go back to a standard screening problem with a size μ_1 of high-value buyers and $1 - \mu_1$ low-value buyers. The **production cost** is assumed to be zero for simplicity. Adding a non-zero constant marginal cost will not

[8] the limit probability of having a buy cascade

change the main results qualitatively. While the seller explicitly sets **the prices only**, the pricing decision implicitly reflects a menu choice, as the seller can exclude a version from the market by making it unreasonably expensive. Regarding the **tie-breaking rule**, we assume agents always buy the product or choose a better version whenever indifferent. This assumption turns out to be without loss given the private belief distribution is atomless (see Footnote 9 and Section 4.2 for detail).

4 Long-Run Demand

This section aims to derive a long-run demand function from limit learning outcomes. We will begin by analyzing how public and private information influences an agent's choice under different prices. The learning dynamics are then characterized by a public likelihood ratio process and the long-run demand is simply the limit probability of a buy cascade for each version.

4.1 Agents' Problem

The expected payoff for agent t, given any prices $\boldsymbol{p}$, public action history h_t and private signal s_t is

$$\mathbb{E}(U(a_t, \boldsymbol{p}, \omega)|h_t, s_t) = \begin{cases} (1-\theta_t)u_{L0} + \theta_t u_{L1} - p_L, & a_t = L \\ (1-\theta_t)u_{H0} + \theta_t u_{H1} - p_H, & a_t = H \\ 0, & a_t = r \end{cases}$$

where $\theta_t := \Pr(\omega = 1|h_t, s_t)$ denotes the posterior belief of agent t after observing previous agents' actions and the private signal.

The seller effectively excludes a version from the market if the version's price is unreasonably high. Specifically when the threshold posterior for agents to choose the premium version over rejection is higher than that of the basic version,

$$\theta_H := \frac{p_H - u_{H0}}{u_{H1} - u_{H0}} \leq \theta_L := \frac{p_L - u_{L0}}{u_{L1} - u_{L0}}$$

, the basic version becomes so expensive that no one would ever consider it.[9] This results in agents behaving as they would in a single-product world.

If $\theta_L < \theta_H \leq 1$, agents may choose any of the three actions as their posterior belief increases. The learning process has a richer set of observable actions. Hence, the public history can convey more information to the subsequent agents. A longer learning phase will merge where the public belief may exhibit more variations over time before it arrives at a cascade.

If $\theta_L \leq 1 < \theta_H$, agents only consider the basic version. Since the marginal production cost is zero, the total surplus from selling the premium version is greater than that from the basic version at any posterior belief θ_t. Giving up the premium version is, therefore, sub-optimal for the seller.

We follow Smith and Sørensen (2000) and use the public likelihood ratio $l_t(h_t) := \frac{\Pr(\omega=0|h_t)}{\Pr(\omega=1|h_t)}$ to describe the learning dynamics. Agent t's optimal strategy is summarized as

$$a^*(l_t, \tau_t) = \begin{cases} r, & \theta_L l_t > \tau_t(1 - p_L + l_t p_L) \\ L, & \theta_L l_t \leq \tau_t(1 - p_L + l_t p_L) \text{ and } \theta_\Delta l_t > \tau_t(1 - \theta_\Delta + l_t\theta_\Delta) \\ H, & \theta_\Delta l_t \leq \tau_t(1 - \theta_\Delta + l_t\theta_\Delta) \end{cases}$$

where $\theta_\Delta := \max\{\frac{p_H - p_L - (u_{H0} - u_{L0})}{u_{H1} - u_{L1} - (u_{H0} - u_{L0})}, \theta_L\}$.[10]

[9] Under the current tie-breaking rule, a price schedule that equalizes the two thresholds incentivizes agents to buy premium products. An alternative tie-breaking rule could direct agents towards the basic version. In such a case, the seller would ideally avoid offering a basic version in the first place. We can modify the model by letting the seller choose which version to offer before choosing the prices. All the results remain unchanged.

[10] Under this definition, $\theta_\Delta = \theta_L$ if and only if $\theta_L \geq \theta_H$. As a result, when $\theta_\Delta = \theta_L$, L will never be chosen and we observe a single-version market. When $\theta_\Delta > \theta_L$, $\theta_H > \theta_\Delta > \theta_L$ and we observe a two-version market.

4.2 Learning Dynamics

Definition (Public Likelihood Ratio Process) The public likelihood ratio process $\langle l_t \rangle_{t=1}^{\infty}$ is a stochastic process with an initial state $l_1 = \frac{1-\mu_1}{\mu_1}$. It evolves according to transition probabilities $\rho(a|\omega, l_t) := \Pr(a_{t+1}|\omega, l_t)$:

$$\rho(H|\omega, l_t) = 1 - F^{\omega}(\frac{l_t \theta_{\Delta}}{1 - \theta_{\Delta} + \theta_{\Delta} l_t})$$
$$\rho(r|\omega, l_t) = F^{\omega}(\frac{l_t \theta_L}{1 - \theta_L + \theta_L l_t})$$
$$\rho(L|\omega, l_t) = \max\{1 - \rho(H|\omega, l_t) - \rho(r|\omega, l_t), 0\}$$

and the continuation function $l_{t+1} = \phi(a_t, l_t) := l_t \frac{\rho(a_t|\omega=0, l_t)}{\rho(a_t|\omega=1, l_t)}$.

The state space of the public likelihood ratio can be divided into cascade sets and learning sets. A cascade occurs at public likelihood ratio l if the agent chooses the same action regardless of her private signal. Otherwise, active learning happens. Let $J_a := \{l \in [0, \infty] | a^*(l, \tau) = a \text{ almost surely.}\}$ be the cascade set for action a.

Notice that tie-breaking rules do not matter in the long run. Since the private belief distribution F^{ω} is atomless, ties occur with measure zero. The boundaries of the cascade and learning sets, which shape the seller's long-run profit function, remain the same as we change the tie-breaking rules.

For ease of notation, I will use the following notations more often. Let $\boldsymbol{\beta} = (\bar{\beta}, \underline{\beta})$, where $\bar{\beta} := \frac{\bar{b}}{1-\bar{b}}$ ($\underline{\beta} := \frac{\underline{b}}{1-\underline{b}}$) are the likelihood ratios of the good state versus bad state under the best (worst) private beliefs. Let $x_i := \frac{1-\theta_i}{\theta_i}$ for each version i and $x_{\Delta} := \frac{1-\theta_{\Delta}}{\theta_{\Delta}}$. They represent the "buy likelihood ratios" and move up as prices drop.[11]

Three learning schemes, defined as types of partitions of the state space of $\langle l_t \rangle_{t=1}^{\infty}$, can emerge under different menus. First, in a single-version

[11] The mapping from $\boldsymbol{p}$ to (x_L, x_{Δ}) are one-to-one.

scheme ($x_\Delta = x_L$), we observe a rejection cascade set $[\bar{\beta}x_H, \infty]$ and a buy cascade set for the premium version (henceforth, premium cascade set) $[0, \underline{\beta}x_H]$. In between, active learning dynamics occur. See Figure 1a.

In a two-version market ($x_\Delta < x_L$), we observe a rejection cascade set $[\bar{\beta}x_L, \infty]$ and a premium cascade set $[0, \underline{\beta}x_\Delta]$. A basic cascade set $[\bar{\beta}x_\Delta, \underline{\beta}x_L]$ will emerge in between if and only if the prices are sufficiently apart ($\bar{\beta}x_\Delta \leq \underline{\beta}x_L$). Thus, two possible learning schemes may arise in this two-version market: a basic-out scheme ($\bar{\beta}x_\Delta > \underline{\beta}x_L$, Figure 1b) and a basic-in scheme ($\bar{\beta}x_\Delta \leq \underline{\beta}x_L$, Figure 1c). The basic version will play rather different roles in the two schemes. In the basic-out scheme, it facilitates learning but only in the short run. In the basic-in scheme, it can guarantee the seller a 'base salary'.

4.3 The Limit Learning Outcomes and Demand

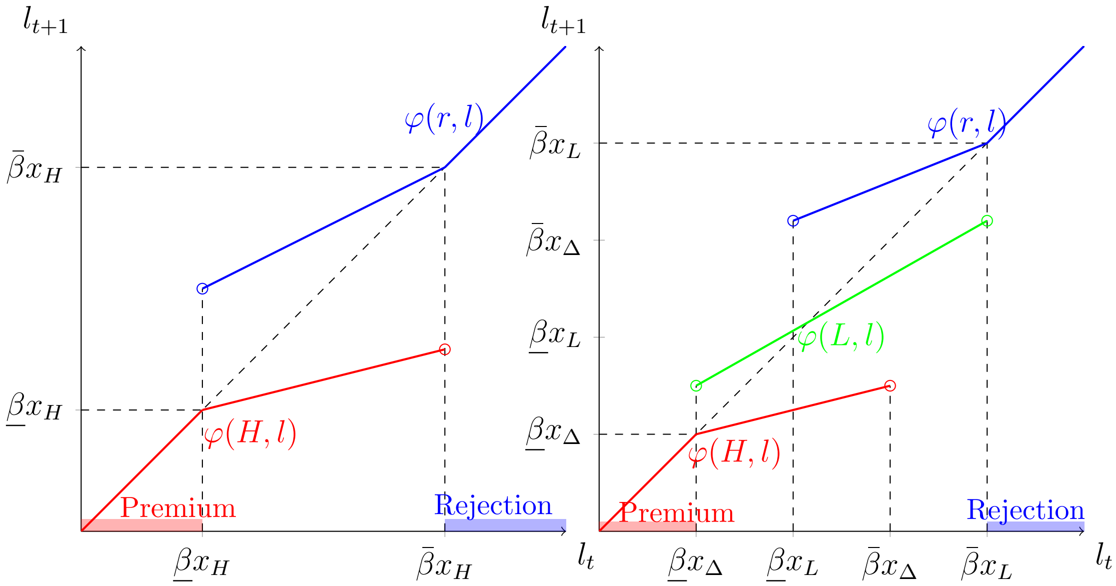
When the seller maximizes the long-run average profit, what happens in finite time does not matter. Lemma 1 shows that the average demand converges to the limit probability of a buy cascade for each version. It provides a closed-form formula for the long-run demand function in each learning scheme. Let $l_\infty = \lim_{t\to\infty} l_t$ be the limit of the public likelihood ratio[12].

Lemma 1. *For any $a \in \{L, H, r\}$,*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} \mathbf{1}(a_s = a) = \Pr(l_\infty \in J_a) \ \ a.s.$$

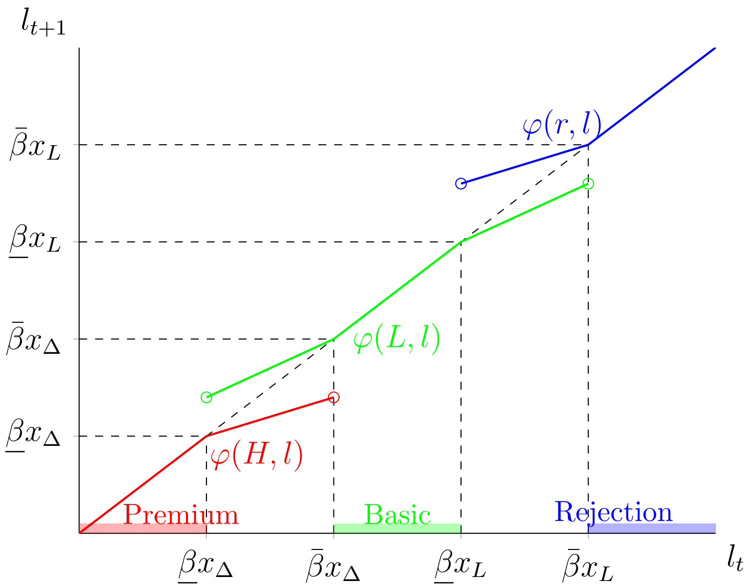
The ex-ante cascade probabilities for each version, denoted by $\lambda(a) := \Pr(l_\infty \in J_a), \forall a \in \{L, H\}$, are

[12] The existence of the limit has been prove by Smith and Sørensen (2000), see Claim 1 below.



(a) A single-version scheme ($x_\Delta = x_L$)

(b) A basic-out scheme ($x_\Delta < x_L, \bar{\beta}x_\Delta > \underline{\beta}x_L$)



(c) A basic-in scheme ($x_\Delta < x_L, \bar{\beta}x_\Delta \leq \underline{\beta}x_L$)

Figure 1: Three types of learning schemes

Note: the three figures plot continuation functions in different learning schemes. In the single-version and basic-out schemes, we run into a premium buy cascade (red shading area) when the public likelihood ratio is low; and a rejection cascade (blue shading area) when the public likelihood ratio is low. If the two prices are far apart, a basic cascade set (green shading area) emerges in between as in the basic-in scheme.

1. $\lambda(H) = \mu_1(1+\underline{\beta}x_H)\frac{\bar{\beta}x_H - l_1}{\bar{\beta}x_H - \underline{\beta}x_H}$ *and* $\lambda(L) = 0$ *in the single-version scheme;*

2. $\lambda(H) = \mu_1(1+\underline{\beta}x_\Delta)\frac{\bar{\beta}x_L - l_1}{\bar{\beta}x_L - \underline{\beta}x_\Delta}$ *and* $\lambda(L) = 0$ *in the basic-out scheme;*

3. $\lambda(H) = \mu_1(1+\underline{\beta}x_\Delta)\frac{\bar{\beta}x_\Delta - l_1}{\bar{\beta}x_\Delta - \underline{\beta}x_\Delta}$ *and* $\lambda(L) = 1 - \lambda(H)$ *in the basic-in scheme if we start from or beyond the basic cascade.*

While the proof of the first part relies heavily on Smith and Sørensen (2000), the second part uses a novel technical result (Lemma 4 in the appendix): when the private belief distribution is differentiable, the public likelihood ratio will hit one of the boundaries of the learning set in the limit. The limit public likelihood ratio never jumps into the interior of the cascade sets. This nice property pins down the support of l_∞. Since the process is a conditional martingale, $E(l_\infty|\omega) = l_1$. From there, we can derive a surprisingly clean expression for the limit cascade probabilities.

5 The Optimal Selling Strategy

In this section, we start by studying the optimal pricing problem in a single-product benchmark. This benchmark analysis highlights the connection between private signal informativeness, learning boundaries and the long-run demand elasticity. We then examine the optimal menu choice and identify new economic forces emerging in the seller's tradeoff between a single- and two-version menu.

5.1 Benchmark: Optimal Pricing with Single Product

This section shows how the seller's pricing incentive changes as signal informativeness improves in a single-product world. It lays the foundation for us to understand when and how observational learning motivates the seller to offer two versions.

Suppose the seller offers a single product at (p, q) with $q \in \mathbb{R}_{++}$ exogenously given. Again, let $u_\omega := u(q, V_\omega), \forall \omega \in \{0, 1\}$ and $\theta_p := \frac{p-u_0}{u_1-u_0}$. Then $x = \frac{1-\theta_p}{\theta_p}$ denotes the buy likelihood ratio.

It is strictly (weakly) optimal to choose a price lower (higher) than a level that already triggers an immediate buy (rejection) cascade. Hence, it suffices to consider the set of x's that satisfy $\underline{\beta}x \leq l_1 \leq \bar{\beta}x$. The single-product pricing problem simplifies to:

$$\begin{aligned}
\max_{x \in [0,\infty)} v_0(x) &:= \mathbb{E}_1(\pi(a, p)) \\
&= \mu_1(1 + \underline{\beta}x)\frac{\bar{\beta}x - l_1}{\bar{\beta}x - \underline{\beta}x}(u_0 + \frac{u_1 - u_0}{1 + x}) \qquad (1) \\
s.t.\ \underline{\beta}x &\leq l_1 \leq \bar{\beta}x
\end{aligned}$$

Similar to the standard monopoly pricing problem, the seller's expected limit average profit consists of two parts: the ex-ante probability of a buy cascade $\lambda(p) = \mu_1(1 + \underline{\beta}x)\frac{\bar{\beta}x - l_1}{\bar{\beta}x - \underline{\beta}x}$ and the price $p = u_0 + \frac{u_1-u_0}{1+x}$. A price increase leads to a lower buy cascade probability (quantity effect) and a higher margin (price effect). The next proposition shows that the quantity effect weakens as the most convincing good and bad news become more informative. The optimal price will increase in general and the learning process starts further away from a buy cascade.

Proposition 1. *1. (Demand Elasticity) The price elasticity of demand $|\frac{d \ln \lambda}{d \ln p}|$ strictly decreases in $\bar{\beta}$ and strictly increases in $\underline{\beta}$.*

2. (Optimality of Immediate Buy Cascade) The Lagrangian multiplier for the constraint $x \leq \frac{l_1}{\underline{\beta}}$ weakly decreases in $\bar{\beta}$ and increases in $\underline{\beta}$.

3. (Optimal Price) The optimal price strictly increases in $\bar{\beta}$ and strictly *decreases in $\underline{\beta}$ at an interior solution.*

The intuition rests on the underlying structure of the long-run demand. A price change affects the demand via shifting the boundaries of the learning set. For instance, a price increase leads to a lower buy likelihood ratio x, and shifts the boundaries downwards relative to the initial belief l_1. The process then starts closer to a rejection cascade, and the long-run buy cascade probability drops.

Signal informativeness determines how responsive the demand is to a price change. The boundaries take a multiplicative form of the likelihood ratio of private belief bounds and the buy likelihood ratio x. Now consider two hypothetical cases with extreme signals. Around the informative limit, $\bar{\beta}$ ($\underline{\beta}$) approaches infinity (zero). A price change barely shifts the learning set. Nor does the relative position of the initial belief change much. Intuitively, too precise information means the seller will need a large price reduction to 'nudge' the same amount of consumers to buy. The quantity effect is thus minimal relative to the price effect.

Conversely, a nearly uninformative signal implies belief updating will be small and the distance between β's shrinks. Even a small price change produces a notable shift of the initial belief's position in the learning set, resulting in a relatively large quantity effect.

A reverse relationship between signal precision and the optimal price arises if an immediate buy cascade is optimal. As the most convincing bad news becomes more informative, the seller must charge a lower price to convince consumers with the worst private signal to buy.

5.2 No Intermediate Cascade

Back to the two-version world, designing the menu in the presence of observational learning is equivalent to choosing a pair of (a) a learning scheme and (b) at which position to start the process.

Some pairs of learning schemes and initial positions can be easily ruled out. For instance, starting below the basic cascade set in the basic-in scheme is strictly dominated by adopting a single-version scheme. The former ends in either a rejection or basic cascade, while the latter results in either a rejection or premium cascade. But a premium cascade is more profitable than a basic cascade as selling a premium product generates a larger total surplus. For the remaining possibilities, we can write the seller's problem as follows:

1. Single-version scheme:

$$\max_{x_H \geq 0} v_H(x_H) := \mu_1(1+\underline{\beta}x_H)\frac{\bar{\beta}x_H - l_1}{\bar{\beta}x_H - \underline{\beta}x_H}(u_{H0} + \frac{\bar{u}_H}{1+x_H}) \tag{2}$$
$$\text{s.t. } x_\Delta = x_L, \underline{\beta}x_H \leq l_1 \leq \bar{\beta}x_H$$

2. Basic-out scheme:

$$\max_{x_L, x_\Delta \geq 0} v(x_L, x_\Delta) := \mu_1(1+\underline{\beta}x_\Delta)\frac{\bar{\beta}x_L - l_1}{\bar{\beta}x_L - \underline{\beta}x_\Delta}(u_{H0} + \frac{\bar{u}_L}{1+x_L} + \frac{\bar{u}_H - \bar{u}_L}{1+x_\Delta})$$
$$\text{s.t. } x_\Delta < x_L, \bar{\beta}x_\Delta > \underline{\beta}x_L \text{ and } \underline{\beta}x_\Delta \leq l_1 \leq \bar{\beta}x_L \tag{3}$$

3. Basic-in scheme (starting between a basic and premium cascade):

$$\max_{x_L, x_\Delta \geq 0} v_L(x_L, x_\Delta) := u_{L0} + \frac{\bar{u}_L}{1+x_L} +$$
$$\mu_1(1+\underline{\beta}x_\Delta)\frac{\bar{\beta}x_\Delta - l_1}{\bar{\beta}x_\Delta - \underline{\beta}x_\Delta}(u_{H0} - u_{L0} + \frac{\bar{u}_H - \bar{u}_L}{1+x_\Delta})$$
$$\text{s.t. } x_\Delta < x_L, \bar{\beta}x_\Delta \leq \underline{\beta}x_L \text{ and } \underline{\beta}x_\Delta \leq l_1 \leq \bar{\beta}x_\Delta \tag{4}$$

Figure 2 plots the constraint sets of the three learning schemes in the buy likelihood ratio space (x_L, x_Δ). $x_\Delta = x_L$ represents the single-version

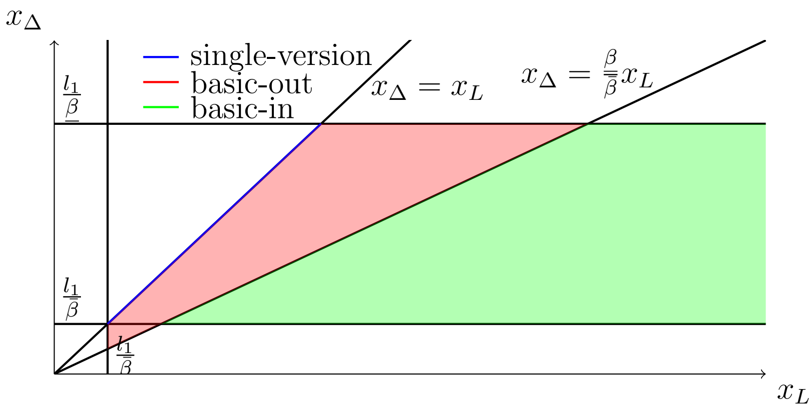


Figure 2: Three learning schemes in (x_L, x_Δ) space.

scheme. $\bar{\beta}x_\Delta = \underline{\beta}x_L$ separates the basic-in and basic-out scheme.

The next result presents two conditions that ensure the suboptimality of a basic cascade.

Lemma 2 (No Intermediate Cascade)**.** *If*

$$u(q_L, V_0) = 0, \ or, \ \frac{\partial}{\partial q_L} \ln \frac{u(q_L, V_0)}{u(q_L, V_1) - u(q_L, V_0)} \geq 0, \tag{5}$$

any prices that implement the basic-in scheme and start the process between a basic and premium cascade are strictly dominated by some prices that implement the single-version scheme.

In words, the second no-intermediate-cascade condition (5) requires that a one percentage increase in the basic quality induces a higher percentage increase in consumers' utility from consuming the basic product in the bad state than their utility difference across states. A large class of utility functions satisfy the condition. For instance, any multiplicatively separable functions $u(\cdot, \cdot)$ fall into this category.

What are the gains and losses when switching from a single-version scheme to a basic-in scheme that starts beyond the basic cascade set? The switch secures a 'base' profit from the basic cascade. However, the seller must concede information rents to future high-belief premium version buyers. The premium price must be sufficiently low to prevent them from deviating to the basic version.

Whenever the condition (5) holds, such a switch is never optimal for any possible $q_L < q_H$. For small enough q_L, the base profit is too low to cover the losses from the information rent. When the basic quality is high enough, the optimal basic-in profit becomes increasing in the basic quality and obtains its upper bound at $q_L = q_H$. But as q_L approaches q_H, the information rent becomes so large that most profits come from the basic version buyers. The seller essentially kicks off the process with an immediate basic cascade. This upper-bound profit, however, can be achieved with an immediate premium cascade in a single-version scheme.

5.3 Existence of an Optimal Menu

The existence of an optimal menu can be problematic with such a fragmented objective function. It turns out that anything within the basic-out region of Figure 2 is either dominated by the boundary above ($x_\Delta = x_L$) or the one below ($x_\Delta = \underline{\beta} x_L / \bar{\beta}$). The latter happens because widening the price gap between versions helps minimize the information rent. But we can never reach the lower constraint. A basic cascade occurs immediately once we push the constraint to equality, causing the profit to jump discontinuously to the basic-in form. It is, however, possible to find an ε-optimal menu.

Theorem 1. *An ε-optimal menu exists. Furthermore, the optimal learning scheme is either a single-version scheme, with a unique optimal menu* *defined by $x_H = x_1^*$ where x_1^* solves the single-version problem*

$$\max_{x\geq 0} v_1(x) = \mu_1(1+\underline{\beta}x)\frac{\bar{\beta}x - l_1}{\bar{\beta}x-\underline{\beta}x}(u_{H0}+\frac{\bar{u}_H}{1+x}) \qquad (6)$$
$$s.t.\ \underline{\beta}x \leq l_1 \leq \bar{\beta}x$$

; or a basic-out scheme, in which case we can find a sequence of ε-optimal menus that converges to $(x_L, x_\Delta) = (x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^*)$ where x_2^* solves the 'limit' profit maximization problem at the binding basic-out constraint:

$$\max_{x\geq 0} v_2(x) = \mu_1(1+\frac{\underline{\beta}^2}{\bar{\beta}}x)\frac{\bar{\beta}x - l_1}{\bar{\beta}x-\frac{\underline{\beta}^2}{\bar{\beta}}x}(u_{H0}+\frac{\bar{u}_L}{1+x}+\frac{\bar{u}_H-\bar{u}_L}{1+\frac{\underline{\beta}}{\bar{\beta}}x}) \qquad (7)$$
$$s.t.\ \frac{\underline{\beta}^2}{\bar{\beta}}x \leq l_1 \leq \bar{\beta}x$$

The objective function in the basic-out scheme is twice continuously differentiable within the basic-out constraint set. Therefore, the comparative statics of the optimal menu structure based on the limit basic-out problem provide a good approximation of the changes in the ε-optimal menus for sufficiently small ε.

5.4 Why Two Versions: The Role of Signal Informativeness

Without any production cost, keeping a premium version in the market is always optimal. The seller's problem then boils down to whether to introduce a cheaper basic version. Let v_1^* be the value function of the single-version problem and v_2^* be the value function of the limit basic-out problem.[13] The first result of this section gives a sufficient and necessary condition for a single-version menu to be optimal for all possible signal

[13] These value functions vary with all the primitives: $\bar{b}, \underline{b}, \mu_1, u, q_H, q_L, V_0, V_1$. We suppress unnecessary parameters to simplify notations throughout the rest of the paper.

precision. If the condition fails, a two-version menu is optimal around the informative limit while a single-version menu remains optimal around the noisy limit.

Lemma 3. *Assume condition (5) holds.* $v_2^*(\bar{b}, \underline{b}) < v_1^*(\bar{b}, \underline{b})$ *for all possible* $\underline{b} \in (0, \frac{1}{2})$ *and* $\bar{b} \in (\frac{1}{2}, 1)$ *if*

$$\mu_1 u(q_H, V_1) \leq u(q_H, V_0). \tag{8}$$

Otherwise, there exists $\delta > 0$ *such that for any* $\bar{b} \in (\frac{1}{\delta}, \infty)$ *and* $\underline{b} \in (0, \delta)$, $v_2^*(\bar{b}, \underline{b}) > v_1^*(\bar{b}, \underline{b})$*; and* $\delta' > 0$ *such that for any* $\bar{b} \in (\frac{1}{2}, \frac{1}{2} + \delta')$ *and* $\underline{b} \in (\frac{1}{2} - \delta', \frac{1}{2})$, $v_2^*(\bar{b}, \underline{b}) < v_1^*(\bar{b}, \underline{b})$.

When the expected surplus from learning the state is limited, excluding the basic option is more profitable regardless of signal informativeness. This is intuitive in the case of fully informative or unbounded private signals. Either we let consumers learn, bet on the prior probability of a good core product, and extract all the surplus from premium version buyers. Or we minimize the risk by inducing an immediate buy 'cascade' with a low price. Comparing profits in the two scenarios delivers the single-version condition (8).

What is less intuitive is why a single-version scheme remains optimal as the private signal becomes less precise. The rationale relates to the seller's incentive in the single-product benchmark. As the signal becomes noisy, long-run demand becomes more responsive to price and a lower price is optimal. As a result, if triggering an immediate buy 'cascade' with a cheap premium version is optimal around the informative limit, the strategy must remain optimal for all other possible signal precisions. This kills the versioning incentive since adding a basic version only reduces the profit by introducing information rents.

Theorem 2 (Versioning)**.** *Assume the no-intermediate-cascade condition (5) holds and the single-version condition (8) fails.*

For any $\underline{b} \in (0, \frac{1}{2})$*, there exists a threshold* $\bar{b}^* \in (\frac{1}{2}, 1)$ *such that* $v_2^*(\bar{b}) \geq v_1^*(\bar{b})$ *if and only if* $\bar{b} \geq \bar{b}^*$*. Likewise, fixing any* $\bar{b} \in (\frac{1}{2}, 1)$*, there exists a threshold* $\underline{b}^* \in (0, \frac{1}{2})$ *such that* $v_2^*(\underline{b}) \geq v_1^*(\underline{b})$ *if and only if* $\underline{b} \leq \underline{b}^*$*.*

Theorem 2 presents a threshold property of how signal informativeness affects the optimal number of menu options. The seller will add a cheaper basic version to the menu if and only if the private signal informativeness exceeds a threshold, in the sense that the most convincing good and bad news is sufficiently informative.

From a technical perspective, discussing how the optimal menu structure changes with signal informativeness will be a non-monotone comparative statics analysis. We want to show that the optimal solution (x_L, x_Δ) moves from the single-version constraint to the basic-out constraint, as the private signal becomes more precise, but never switches back. It is difficult to directly compare value functions in the two single-variable problems (6) and (7), due to the lack of explicit solutions and differing constraint sets. Instead, we aim to prove that the first-order partial derivative of the basic-out value function with respect to $\bar{\beta}$ or $\underline{\beta}$ exceeds that of the single-version value function, whenever the two value functions are equal.

The key step is to introduce a modified basic-out problem. Consider a marginal increase in $\bar{\beta}$. Instead of letting $\bar{\beta}$ change in both the profit function (direct effect) and through the binding basic-out constraint (indirect effect), I shut down the indirect effect by fixing the $\bar{\beta}$'s that enter into the profit function in (7) via the basic-out constraint at a level that equates the two value functions. This indirect effect expands the basic-out constraint set, and thus killing the effect never works against our goal. So we can work with the modified value function, which significantly simplifies the problem.

Intuitions. Why is it optimal to add a cheap basic version only when the private signal is informative?

On the cost side, a cheaper basic version forces the seller to leave information rents to premium version buyers. The gap in the highest possible premium price in the single-version scheme (6) and the basic-out scheme (7) demonstrates this information rent: $p_H^1(\frac{l_1}{\beta}) - p_H^2(\frac{l_1}{\beta}) = (\bar{u}_H - \bar{u}_L)(\frac{1}{1+\frac{l_1}{\beta}} - \frac{1}{1+\frac{l_1}{\underline{\beta}}}) > 0$. This rent increases as the private signal becomes more informative.

On the benefit side, a cheap basic version relaxes the 'no-immediate-rejection-cascade' constraint, allowing the seller to charge a high premium price that would have caused an immediate rejection cascade if she offered a single premium version. A two-version menu essentially stretches the learning set towards lower beliefs so the first several low-belief agents can try the new product at a low cost. The seller then extracts more surplus from high-belief buyers by lifting the premium version price without advancing the onset of a rejection cascade.

This role of the basic version works, however, only when the private signal is sufficiently informative. An informative signal weakens the quantity effect and incentivizes the seller to lift the premium version price rather than lower it to hedge against the rejection cascade risk. With a noisy signal, however, a strong quantity effect makes the latter more profitable. The seller prefers a low price in the first place. Hence, the basic version does not work. Intuitively, there's no point in providing a cheaper option if the seller already sells an affordable premium product.

An alternative perspective to look at the result is that adding a cheap basic version rotates the long-run demand function clockwise. While I cannot show the premium price as a function of the long-run demand λ rotates, Lemma 5 proves a single-crossing property of 'pseudo prices' as functions of the conditional probability $\lambda(H|\omega = 1)$: The pseudo premium price function of a two-version (basic-out) menu crosses that of a single-version menu only once and from above. Hence, the optimal two-version menu outperforms the optimal single-version menu only when the 'price' effect is much stronger than the 'quantity' effect.[14]

6 Concluding Remarks

This paper investigates the optimality of a multi-version menu in new product markets with observational learning. In such a scenario, a cheaper basic version can take two possible roles. First, it can secure a 'base' profit if the menu implements a learning process ending with either a basic cascade or premium cascade. Second, the basic version can facilitate social learning in the beginning but disappears in the long run. It postpones the onset of a rejection cascade and allows the seller to charge a higher price over the premium version. So she can reap the fruit of learning when the market recognizes its value. Lemma 2 shows the first role does not work even if the seller cares about long-run profit.

Another unique insight of the paper is that private signal informativeness affects the price elasticity of long-run demand, leading to qualitatively different selling strategies. In a market with noisy private information, the seller offers a single cheap premium version. When consumers arrive with precise private information, the second role of the basic version works and adding it to the menu becomes optimal.

From an applied perspective, the conditions identified in Lemma 2, Lemma 3, and Theorem 2 allow us to obtain sharp predictions over when we will see more versioning of the kind that aims to facilitate observational learning, in addition to those aiming to screen different types of consumers.

[14] Johnson and Myatt (2006) has a very nice discussion on how demand rotations can lead to versioning. The interpretation here has a subtle difference: adding a version leads to demand rotation rather than the other way around.

Appendices

Appendix A Proofs

A.1 Proof of Lemma 1

To start with, let me present some results on belief convergence and action convergence. Most of them have been proved in Smith and Sørensen (2000). Then I will prove Lemma 4, which characterises the support of the limit public likelihood ratio. As a final step, I will show how to derive the limit cascade probabilities.

Claim 1. *Conditional on state $\omega = 1$, the public likelihood process $\langle l_t \rangle_{t=1}^{\infty}$ is a martingale. Assume $\omega = 1$. The public likelihood ratio converges almost surely to a random variable $l_\infty = \lim_{t\to\infty} l_t \in [0, \infty)$. Also, a cascade will happen in the limit.*

Likewise, conditional on state $\omega = 0$, the process of the reciprocal of $\langle l_t \rangle_{t=1}^{\infty}$ is a martingale. Assume $\omega = 0$. The inverse public likelihood ratio converges almost surely to a random variable $\tilde{l}_\infty \in [0, \infty)$. A cascade will happen in the limit.

The first part of the claim has been proved in Smith and Sørensen (2000) (see their Lemma 3 and Theorem 1 (a)). The proof for the second part follows suit by replacing l_t with $\tilde{l}_t := 1/l_t$ and conditioning on the bad state $\omega = 0$. The inverse public likelihood ratio process is formally defined by: (1) an initial state $\tilde{l}_1 = \frac{\mu_1}{1-\mu_1}$ and (2) transition probabilities

$\rho(a|\omega, \tilde{l}_t) := \Pr(a_{t+1}|\omega, \tilde{l}_t)$:

$$\rho(H|\omega, \tilde{l}_t) = 1 - F^{\omega}(\frac{\theta_\Delta}{\tilde{l}_t(1-\theta_\Delta)+\theta_\Delta})$$
$$\rho(r|\omega, \tilde{l}_t) = F^{\omega}(\frac{\theta_L}{\tilde{l}_t(1-\theta_L)+\theta_L})$$
$$\rho(L|\omega, \tilde{l}_t) = \max\{1-\rho(H|\omega, \tilde{l}_t) - \rho(r|\omega, \tilde{l}_t), 0\}$$

and the continuation function $\tilde{l}_{t+1} = \phi(a_t, \tilde{l}_t) := \tilde{l}_t \frac{\rho(a_t|\omega=1,\tilde{l}_t)}{\rho(a_t|\omega=0,\tilde{l}_t)}$.

The corollary of Lemma 3 in Smith and Sørensen (2000) shows action convergence almost surely obtains if the private belief distribution is atomless. Combined with the claim above, it is easy to see that $\frac{1}{t}\sum_{s=1}^{t} \mathbf{1}(a_s = a)$ converges almost surely to $\Pr(l_\infty \in J_a)$.

The next Lemma shows the public likelihood ratio will almost surely hit one of the two boundaries of the learning set in the limit.

Lemma 4. *The limit public likelihood ratio has a binary support if F^ω is differentiable. $supp(l_\infty) = \{l_1, l_2\}$ where l_1 and l_2 are the boundaries of the learning set from which we start. Similarly, $supp(\tilde{l}_\infty) = \{\tilde{l}_1, \tilde{l}_2\}$ where $\tilde{l}_1$ and $\tilde{l}_2$ are the boundaries of the learning set from which we start.*

Proof. Let's consider the public likelihood ratio process $\langle l_t \rangle_{t=1}^{\infty}$. The proof for the reciprocal process $\langle \tilde{l}_t \rangle_{t=1}^{\infty}$ is analogous. I will first prove the statement is true with two actions and then move to the case of three actions. As you will see, it is easy to extend the result to the case with any finite number of actions. The key idea is to show that the continuation function $\phi(a, l)$, as defined in 4.2, never jumps out of the boundaries when we start from within the learning set. Given any $\theta \in [0, 1]$, let $\tau(l) = \frac{l\theta}{1-\theta+l\theta}$ represent a threshold private belief in the agent's optimal strategy. Denote the two actions by a_1 and a_2. The continuation function for each action and the boundary likelihood ratios of its associated cascade set must satisfy

$$\phi(a_1, l) = l\frac{1 - F^0(\tau(l))}{1 - F^1(\tau(l))}, \phi(a_2, l) = l\frac{F^0(\tau(l))}{F^1(\tau(l))}$$
$$l_1 = \phi(a_1, l_1) = \bar{\beta}\frac{1-\theta}{\theta}, l_2 = \phi(a_2, l_2) = \underline{\beta}\frac{1-\theta}{\theta}$$

We want to show (1) $\phi(a_1, l) \geq l_1, \forall\ l \in (l_1, l_2)$ and (2) $\phi(a_2, l) \leq l_2, \forall\ l \in (l_1, l_2)$. I will only prove the first statement here. A similar argument works for the second. Suppose by contradiction $\phi(a_1, l) < l_1$ for some $l \in (l_1, l_2)$. Since the private belief distributions F^ω is atomless, $\phi(a_1, \cdot)$ is continuous on $[l_1, l_2]$. So we can find some $l^* \in (l_1, l_2)$ such that $\frac{\partial\phi(a_1, l^*)}{\partial l} < 0$ and $\phi(a_1, l^*) = l_1$. The following analysis aims to show $\frac{\partial\phi(a_1, l^*)}{\partial l} \geq 0$, which then leads to a contradiction.

By definition of ϕ, $\phi(a_1, l^*) = l_1$ implies $l^*\frac{1-F^0(\tau(l^*))}{1-F^1(\tau(l^*))} = l_1$. Let $\tau^* := \tau(l^*)$. We have

$$\frac{f^0(\tau^*)}{f^1(\tau^*)} = \frac{1-\tau^*}{\tau^*} = \frac{l_1}{l^*\bar{\beta}} = \frac{1-F^0(\tau^*)}{1-F^1(\tau^*)}\frac{1}{\bar{\beta}}.$$

The first equation follows from Lemma A.1 (a) in Smith and Sørensen (2000). The second follows from the definition of $\tau(l)$ and l_1. Rearrange the equation and we have $\frac{f^0(\tau^*)}{1-F^0(\tau^*)} = \frac{f^1(\tau^*)}{1-F^1(\tau^*)}\frac{1}{\bar{\beta}}$.

$$\begin{aligned}
\frac{\partial\phi(a_1, l^*)}{\partial l} &= \frac{(1-F^0(\tau^*))}{(1-F^1(\tau^*))}[1 + \tau^*(1-\tau^*)(\frac{f^1(\tau^*)}{1-F^1(\tau^*)} - \frac{f^0(\tau^*)}{1-F^0(\tau^*)})] \\
&= \frac{(1-F^0(\tau^*))}{(1-F^1(\tau^*))}[1 + \tau^*(1-\tau^*)(\bar{\beta}-1)\frac{f^0(\tau^*)}{1-F^0(\tau^*)})] \\
&\geq 0. \text{ (because } \bar{\beta} > 1)
\end{aligned}$$

Now let's consider the case of three actions $\{a_1, a_2, a_3\}$. Assume we have cascade sets for a_1 and a_3. Let the boundaries of the learning set be $l_1 = \phi(a_1, l)$ and $l_3 = \phi(a_3, l) > l_1$. Suppose also $1 > \tau_1 := \frac{l\theta_1}{1-\theta_1+l\theta_1} >$

$\tau_2 := \frac{l\theta_2}{1-\theta_2+l\theta_2} > 0$ for some $(\theta_1, \theta_2) \in [0,1]^2$. An agent chooses a_1 if they receive a private belief $\tau \geq \tau_1$; a_3 if $\tau < \tau_2$; a_2 if $\tau_2 \leq \tau < \tau_1$. Then the continuation functions are

$$\phi(a_1, l) = l\frac{1 - F^0(\tau_1(l))}{1 - F^1(\tau_1(l))}$$
$$\phi(a_2, l) = l\frac{F^0(\tau_1(l)) - F^0(\tau_2(l))}{F^1(\tau_1(l)) - F^1(\tau_2(l))}$$
$$\phi(a_3, l) = l\frac{F^0(\tau_2(l))}{F^1(\tau_2(l))}$$

Repeat the same steps as in the above two-action proof and we have, for all $l \in (l_1, l_3)$, (1) $\phi(a_1, l) \in (l_1, l_3)$ and (2) $\phi(a_3, l) \in (l_1, l_3)$. Intuitively, a belief update following the middle action should be smaller than that from extreme actions. The last part of the proof shows $\phi(a_1, l) < \phi(a_2, l) < \phi(a_3, l)$ whenever the continuation functions are well-defined at $l \in (l_1, l_3)$[15].

We can prove a slightly more general statement:

$$\frac{F^0(\tau'') - F^0(\tau')}{F^1(\tau'') - F^1(\tau')} < \frac{F^0(\tau') - F^0(\tau)}{F^1(\tau') - F^1(\tau)}, \forall 0 \leq \tau < \tau' < \tau'' \leq 1. \tag{9}$$

Because $\frac{f^0(\tau)}{f^1(\tau)} = \frac{1-\tau}{\tau}$, we have $F^0(\tau'') - F^0(\tau') = \int_{\tau'}^{\tau''} \frac{1-\tau}{\tau} dF^1 < \frac{1-\tau'}{\tau'}(F^1(\tau'') - F^1(\tau'))$. Similarly, $F^0(\tau') - F^0(\tau) > \frac{1-\tau'}{\tau'}(F^1(\tau') - F^1(\tau))$. Hence, the left-hand side of equation 9 $< \frac{1-\tau'}{\tau'} <$ its right-hand side. □

Finally, we derive the limit cascade probabilities. As an illustration, I will present how to derive the ex-ante cascade probabilities in the single-version scheme. The other cases are similar.

Let $l_r = \bar{\beta} x_H$ and $l_H = \underline{\beta} x_H$ denote the boundaries of the learning set for the public likelihood ratio process and $1/l_r$ and $1/l_H$ will be those for the inverse public likelihood ratio process. By the Dominated Convergence Theorem, the conditional premium cascade probabilities $\lambda(H|\omega) := \Pr(l_\infty \in J_H)$

[15] That is, when $\rho(a|\omega, l) > 0$.

satisfy

$$E(l_\infty|\omega=1) = \lambda(H|\omega=1)l_H + (1-\lambda(H|\omega=1))l_r = l_1$$
$$E(\tilde{l}_\infty|\omega=0) = \lambda(H|\omega=0)\frac{1}{l_H} + (1-\lambda(H|\omega=0))\frac{1}{l_r} = \frac{1}{l_1}.$$

Hence, the ex-ante premium cascade probability is

$$\lambda(H) = \mu_1\lambda(H|\omega=1) + (1-\mu_1)\lambda(H|\omega=0) = \mu_1(1+\underline{\beta}x_H)\frac{\bar{\beta}x_H - l_1}{\bar{\beta}x_H - \underline{\beta}x_H}.$$

A.2 Proof of Proposition 1

Following Lemma 1, it is easy to establish the seller's problem as in 1.

Note that $\ln\lambda = \ln\mu_1 + \ln(1+\underline{\beta}x) + \ln(\bar{\beta}x - l_1) - \ln(\bar{\beta}-\underline{\beta}) - \ln x$ and $\ln p = \ln(u_0 + \frac{u_1-u_0}{1+x})$. Thus, we have

$$\begin{aligned}
|\frac{d\ln\lambda}{d\ln p}| &= |\frac{(\frac{\underline{\beta}}{1+\underline{\beta}x} + \frac{\bar{\beta}}{\bar{\beta}x-l_1} - \frac{1}{x})dx}{-\frac{u_1-u_0}{(1+x)^2}\frac{1}{u_0+(u_1-u_0)/(1+x)}dx}| \\
&= \frac{1}{u_1-u_0}(1+x)(u_0x+u_1)[\frac{1}{\frac{1}{\underline{\beta}}+x} + \frac{1}{x-\frac{l_1}{\bar{\beta}}} - \frac{1}{x}]. \qquad (10)
\end{aligned}$$

It's easy to see $|\frac{d\ln\lambda}{d\ln p}|$ strictly decreases (increases) in $\bar{\beta}$ ($\underline{\beta}$).

Now let us look at the optimal price in problem 1. Since the objective function is continuous and nonnegative on the feasible set, it is without loss to work with $\ln v_0$. The first-order derivative of $\ln v_0$ with respect to x is

$$\frac{\partial \ln v_0}{\partial x} = \frac{\underline{\beta}}{1+\underline{\beta}x} + \frac{\bar{\beta}}{\bar{\beta}x - l_1} - \frac{1}{x} - \frac{u_1-u_0}{(1+x)(u_0x+u_1)} \qquad (11)$$

Note that $\frac{\partial\ln v_0}{\partial x}$ approaches infinity as $x \to \frac{l_1}{\bar{\beta}} + 0$ (close to an immediate rejection cascade). Thus, $x \geq \frac{l_1}{\bar{\beta}}$ never binds.

Fix any $\underline{\beta} \in (0,1)$. $\frac{\partial\ln v_0}{\partial x}$ strictly decreases in $\bar{\beta}$ for all $x > \frac{l_1}{\bar{\beta}}$. The optimal buy likelihood ratio x^* must be weakly decreasing in $\bar{\beta}$ (strictly so if we have an interior solution). Equivalently, the optimal price goes up as $\bar{\beta}$ increases.

Fix any $\bar{\beta} \in (1, \infty)$. $\frac{\partial \ln v_0}{\partial x}$ strictly increases in $\underline{\beta}$ for all $x > 0$. It has two implications. First, the interior solution strictly increases in $\underline{\beta}$. Second, it is more likely to hit the boundary $x \leq \frac{l_1}{\underline{\beta}}$. When that happens, $x^* = \frac{l_1}{\underline{\beta}}$ strictly decreases in $\underline{\beta}$. Therefore, as $\underline{\beta}$ decreases from 1 to 0, the optimal buy likelihood ratio x^* increases (or the price drops) until the seller finds it optimal not to trigger an immediate premium cascade. From then on, the optimal buy likelihood ratio decreases and the optimal price goes up.

A.3 Proof of Lemma 2

Notice that the basic-in constraint ($\bar{\beta} x_\Delta \leq \underline{\beta} x_L$) binds in the basic-in problem (4). Otherwise, the seller can always increase her payoff by decreasing x_L without violating any other constraints. Plug the constraint into the objective function and we have the following optimization problem:

$$\max_x v_L(\frac{\bar{\beta}}{\underline{\beta}}x, x; q_L) \text{ subject to } \frac{l_1}{\bar{\beta}} \leq x \leq \frac{l_1}{\underline{\beta}} \tag{12}$$

Let $x^*(q_L)$ denote the solution to this program and v_L^* its value function. v_H^* denotes the value function for problem (2).

Define the difference between the basic-in profit $v_L(\frac{\bar{\beta}}{\underline{\beta}}x, x; q_L)$ and the single-version profit $v_H(x)$ as

$$\begin{aligned} D(x, q_L) &:= v_L(\frac{\bar{\beta}}{\underline{\beta}}x, x; q_L) - v_H(x) \qquad (13) \\ &= u_{L0}[1 - \mu_1(1 + \underline{\beta}x)\frac{\bar{\beta}x - l_1}{\bar{\beta}x - \underline{\beta}x}] + \bar{u}_L[\frac{1}{1 + \frac{\bar{\beta}}{\underline{\beta}}x} - \mu_1(1 + \underline{\beta}x)\frac{\bar{\beta}x - l_1}{\bar{\beta}x - \underline{\beta}x}\frac{1}{1 + x}] \end{aligned}$$

for $x \in [\frac{l_1}{\bar{\beta}}, \frac{l_1}{\underline{\beta}}]$.

Here are some preliminary properties of D that I will use in the following steps. Let $m_0(x) := 1 - \mu_1(1+\underline{\beta}x)\frac{\bar{\beta}x-l_1}{\bar{\beta}x-\underline{\beta}x}$ and $m_1(x) := \frac{1}{1+\frac{\bar{\beta}}{\underline{\beta}}x} - \mu_1(1+\underline{\beta}x)\frac{\bar{\beta}x-l_1}{\bar{\beta}x-\underline{\beta}x}\frac{1}{1+x}$. So $D(x,q_L) = u_{L0}m_0(x) + \bar{u}_L m_1(x)$.

1. It it easy to show that $m_0(x) \geq 0$, $\frac{\partial m_0}{\partial x} < 0$ and $\frac{\partial^2 m_0}{\partial x^2} > 0$ for $x \in [\frac{l_1}{\bar{\beta}}, \frac{l_1}{\underline{\beta}}]$.

2. Note also that $m_1(x) = \frac{1}{1+x}[\frac{1+x}{1+\frac{\bar{\beta}}{\underline{\beta}}x} - \mu_1(1+\underline{\beta}x)\frac{\bar{\beta}x-l_1}{\bar{\beta}x-\underline{\beta}x}]$. Inside the bracket is a function strictly decreasing in x. Hence, $\frac{\partial m_1}{\partial x} < 0$ whenever $m_1(x) \geq 0$.

The main proof assumes the second condition ($\frac{\partial}{\partial q_L}\ln\frac{u(q_L,V_0)}{u(q_L,V_1)-u(q_L,V_0)} \geq 0$) holds. Then it will be rather straightforward why $u(q_L,V_0) = 0$ also works. I will briefly comment on this at the end of the proof.

Step 1: $D(x,q_L) > 0 \Rightarrow \frac{\partial D}{\partial q_L} = \frac{\partial v_L}{\partial q_L} > 0$ for all $(x,q_L) \in \mathbb{R}^2_{++}$.

Note first that $D(x,q_L) > 0 \Leftrightarrow m_1(x) \geq \frac{-u_{L0}m_0(x)}{\bar{u}_L}$.

$$\frac{\partial D}{\partial q_L} = \frac{\partial v_L}{\partial q_L}(\frac{\bar{\beta}}{\underline{\beta}}x, x; q_L)) = \frac{\partial u_{L0}}{\partial q_L}m_0 + \frac{\partial \bar{u}_L}{\partial q_L}m_1$$
$$> \frac{\partial u_{L0}}{\partial q_L}m_0 - \frac{\partial \bar{u}_L}{\partial q_L}\frac{u_{L0}m_0}{\bar{u}_L} = m_0 u_{L0}[\frac{\frac{\partial u_{L0}}{\partial q_L}}{u_{L0}} - \frac{\frac{\partial \bar{u}_L}{\partial q_L}}{\bar{u}_L}] = m_0 u_{L0}\frac{\partial}{\partial q_L}\ln\frac{u_{L0}}{\bar{u}_L} \geq 0$$

The last inequality follows from our second no-intermediate-cascade condition.

Step 2: $\frac{\partial D}{\partial x}(x,q_L) \geq 0 \Rightarrow D(x,q_l) < 0$ for all $(x,q_L) \in [\frac{l_1}{\bar{\beta}}, \frac{l_1}{\underline{\beta}}] \times [0,q_H]$.

Fix any $q_L \in [0,q_H]$. To begin with, I would like to show $m_1'(x) > 0 \Rightarrow m_1''(x) > 0$. Suppose $m_1'(x) > 0$. Consider $\ln(m_1'(x)) = -\ln\frac{\bar{\beta}}{\underline{\beta}} + 2\ln(1+\frac{\bar{\beta}}{\underline{\beta}}x) - \ln(\bar{\beta}\underline{\beta} + \frac{l_1}{x^2}) + \ln(1+\underline{\beta}x) + \ln(\bar{\beta}x - l_1) - \ln x - \ln(1+x)$. Then, $\frac{\partial}{\partial x}[\ln(m_1'(x))] = \frac{2}{\frac{\underline{\beta}}{\bar{\beta}}+x} + \frac{2l_1x^{-3}}{\bar{\beta}\underline{\beta}+\frac{l_1}{x^2}} + \frac{\underline{\beta}}{1+\underline{\beta}x} + \frac{\bar{\beta}}{\bar{\beta}x-l_1} - \frac{1}{x} - \frac{1}{1+x} > 0$ because the first term $\frac{2}{\frac{\underline{\beta}}{\bar{\beta}}+x}$ is obviously higher than $\frac{1}{1+x}$ and the third term $\frac{\bar{\beta}}{\bar{\beta}x-l_1} > \frac{1}{x}$. It follows that $m_1''(x) > 0$ whenever $m_1'(x) > 0$.

As a result, whenever $\frac{\partial}{\partial x}D(x, q_L) = u_L m_0'(x) + \bar{u}_L m_1'(x) > 0$, given the properties of m_0, we must have $m_1'(x) > 0$ and thus $\frac{\partial^2 D}{\partial x^2}(x, q_L) = u_L m_0''(x) + \bar{u}_L m_1''(x) > 0$. It implies that $\frac{\partial D}{\partial x}$ keeps strictly increasing after it turns positive for the first time. Let $\hat{x} := \inf\{x : \frac{\partial D}{\partial x} \geq 0\}$. If $x \in [\hat{x}, \frac{l_1}{\underline{\beta}}]$, $D(x, q_L) < D(\frac{l_1}{\underline{\beta}}) = \bar{u}_L m_1(x)$. But $\frac{\partial D}{\partial x} \geq 0 \Rightarrow m_1'(x) > 0$ (the first preliminary property 1) and hence $m_1(x) < 0$ (following the premilinary property 2). So $D(x, q_L) < 0$ for all $x \in [\hat{x}, \frac{l_1}{\underline{\beta}}]$.

Step 3: There exists $\hat{q}_L \in [0, q_H]$ such that (1) for all $q_L \in (0, \hat{q}_L)$, $D(x^*(q_L), q_L) < 0$; (2) for all $q_L \in (\hat{q}_L, q_H)$, $v_L(x^*(q_L), q_L) \leq v_L(x^*(q_H), q_H) \leq v_H(\frac{l_1}{\underline{\beta}})$.

Let us start with proving the following statement: $D(x^*(q_L), q_L) \geq 0 \Rightarrow \frac{dD}{dq_L}(x^*(q_L), q_L) = \frac{\partial D}{\partial q_L}(x^*(q_L), q_L) + \frac{\partial D}{\partial x}(x^*(q_L), q_L)\frac{dx^*}{dq_L}(q_L) > 0$. The contrapositive of the result in step 2 tells us $\frac{\partial D}{\partial x} < 0$ when $D \geq 0$. For any (x, q_L) such that $D(x, q_L) \geq 0$,

$$\begin{aligned}\frac{\partial^2 v_L}{\partial q_L \partial x}(x, q_L) &= \frac{\partial u_{L0}}{\partial q_L}(q_L)m_0'(x) + \frac{\partial \bar{u}_L}{\partial q_L}(q_L)m_1'(x) \\ &< \frac{\frac{\partial u_{L0}}{\partial q_L}}{u_{L0}}(u_{L0}m_0'(x) + \bar{u}_L m_1'(x)) = \frac{\frac{\partial u_{L0}}{\partial q_L}}{u_{L0}}\frac{\partial D}{\partial x} < 0.\end{aligned}$$

The first inequality follows from the second no-intermediate-cascade condition. The second inequality follows from the assumption that u is strictly increasing. Also, when $D(x^*(q_L), q_L) > 0$, all the constraints are slack except the basic-in constraint in problem (12). By the Implicit Function Theorem, $\frac{dx^*}{dq_L}(q_L) < 0$. Combining with step 1 we know $\frac{dD}{dq_L}(x^*(q_L), q_L) = \frac{\partial D}{\partial q_L} + \frac{\partial D}{\partial x}\frac{dx^*}{dq_L} > 0$ whenever $D(x^*(q_L), q_L) \geq 0$.

It follows that there exists $\hat{q}_L \in [0, q_H]$. For any $q_L \in (0, \hat{q}_L)$, $D(x^*(q_L), q_L) < 0$. So $v_L^*(q_L) < v_H(x^*(q_L)) \leq v_H^*$. For any $q_L \in [\hat{q}_L, q_H)$, $D(x^*(q_L), q_L) \geq 0$. By the Envelope theorem, $\frac{dv_L^*}{dq_L}(q_L) = \frac{\partial v_L}{\partial q_L}(x^*(q_L), q_L) = \frac{\partial D}{\partial q_L}(x^*(q_L), q_L) > 0$.

Hence, $v_L^*(q_L) < v_L^*(q_H) = u_{H0} + \frac{\bar{u}_H}{1+\frac{l_1}{\underline{\beta}}} = v_H(\frac{l_1}{\underline{\beta}}) \leq v_H^*$.

Now consider the first no-intermediate cascade condition $u(q_L, V_0) = 0$. In this case, $D(x, q_L) = \bar{u}_L m_1(x)$. Step 1 immediately follows because $\bar{u}_L > 0$ and $\frac{\partial \bar{u}_L}{\partial q_L} > 0$. Step 2 is simply a contrapositive of the preliminary property 1: $\frac{\partial m_1}{\partial x} < 0$ whenever $m_1(x) \geq 0$. A similar argument as in Step 3 brings us to the same conclusion.

A.4 Proof of Theorem 1

Now that we rule out the basic-in scheme, we can merge the single-version and basic-out problem into a general problem

$$\max_{x_L, x_\Delta \geq 0} v(x_L, x_\Delta) := \mu_1(1 + \underline{\beta} x_\Delta) \frac{\bar{\beta} x_L - l_1}{\bar{\beta} x_L - \underline{\beta} x_\Delta}[u_{H0} + \frac{\bar{u}_L}{1 + x_L} + \frac{\bar{u}_H - \bar{u}_L}{1 + x_\Delta}] \tag{14}$$

$$\text{s.t. } x_\Delta \leq x_L, \bar{\beta} x_\Delta > \underline{\beta} x_L \text{ and } \underline{\beta} x_\Delta \leq l_1 \leq \bar{\beta} x_L$$

Denote the constraint set by $\mathscr{C}(\bar{\beta}, \underline{\beta})$ and $\boldsymbol{x} := (x_L, x_\Delta)$. The first-order partial derivative of v with respect to x_Δ is $\frac{\partial v}{\partial x_\Delta}(x_L, x_\Delta) =$

$$\begin{aligned} &\frac{\mu_1(\bar{\beta} x_L - l_1)}{(\bar{\beta} x_L - \underline{\beta} x_\Delta)^2 (1 + x_\Delta)^2}[((u_{H0} + \frac{\bar{u}_L}{1 + x_L})(\bar{\beta}\underline{\beta} x_L + \underline{\beta}) + (\bar{u}_H - \bar{u}_L)\underline{\beta}^2) x_\Delta^2 \\ &+ 2((u_{H0} + \frac{\bar{u}_L}{1 + x_L})(\bar{\beta}\underline{\beta} x_L + \underline{\beta}) + (\bar{u}_H - \bar{u}_L)\underline{\beta}) x_\Delta \\ &+ (u_{H0} + \frac{\bar{u}_L}{1 + x_L})(\bar{\beta}\underline{\beta} x_L + \underline{\beta}) + (\bar{u}_H - \bar{u}_L)((\underline{\beta} - 1)\bar{\beta} x_L + \underline{\beta})] \end{aligned} \tag{15}$$

Fixing any $x_L > 0$, v either increases in x_Δ on $(0, \infty)$ or decreases first and then increases in x_Δ on $(0, \infty)$. Hence, at the optimal solution, x_Δ always hits or stays as close as possible to one of the constraints. If the basic-out scheme outperforms the single-version scheme, it is optimal to stay as close to the basic-out constraint $\bar{\beta} x_\Delta > \underline{\beta} x_L$ as possible. While we cannot reach the constraint, an ε-optimal menu always exists. Remember that x_2^*, as defined in (7), maximises the seller's profit when the basic-out constraint binds[16]. By the definition of the first-order partial derivative, for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that

$$|\frac{v(x_2^*, x_\Delta) - v(x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^*)}{x_\Delta - \frac{\underline{\beta}}{\bar{\beta}}x_2^*} - \frac{\partial v}{\partial x_\Delta}(x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^*)| < \varepsilon, \forall x_\Delta \in (\frac{\underline{\beta}}{\bar{\beta}}x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^* + \delta(\varepsilon)) \tag{16}$$

which implies

$$|v(x_2^*, x_\Delta) - v(x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^*)| < (\varepsilon + |\frac{\partial v}{\partial x_\Delta}(x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^*)|)|x_\Delta - \frac{\underline{\beta}}{\bar{\beta}}x_2^*|, \forall x_\Delta \in (\frac{\underline{\beta}}{\bar{\beta}}x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^* + \delta(\varepsilon)).$$

Hence, $(x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^* + d)$ where $d = \min\{\delta(\varepsilon), \frac{1}{2}, \frac{\varepsilon}{2|\frac{\partial v}{\partial x_\Delta}(x_2^*, \frac{\underline{\beta}}{\bar{\beta}}x_2^*)|}\}$ is an ε-optimal menu[17]. If the single-version scheme ($x_\Delta = x_L$) outcompetes the basic-out scheme, an optimal menu always exists as in (6).

A.5 Proof of Lemma 3

Note first that in the informative limit ($\bar{\beta} \to \infty, \underline{\beta} \to 0$), the conditional probabilities of buy cascades are not well-defined. What I discuss here only informs us of what happens when the private belief bounds get close to the limit.

The outline of proof is as follows. The first step provides preliminary results on the differentiability of v_i^* in $\boldsymbol{\beta}$. Then, I show that a single-version strategy is optimal around the noisy limit. The third step proves that a single-version strategy is optimal for all possible $\boldsymbol{\beta}$ when the single-version condition (8) holds. I finish the proof by comparing the optimal profits from the two versioning strategies around the informative limit.

[16]The solution to (7) must exist because it has a continuous objective function and a bounded, convex feasible set.

[17]d is well-defined since $\frac{\partial v}{\partial x_\Delta}(x_L, x_\Delta)$ never approaches infinity on the bounded feasible set of (x_L, x_Δ) in problem (14)

The first-order derivatives of v_i's w.r.t. x (v_i defined in 6 and 7) will be important to the following analysis:

$$\frac{\partial v_1}{\partial x} = \frac{\mu_1}{1-\frac{\underline{\beta}}{\bar{\beta}}}\left[\left(\underline{\beta}+\frac{l_1}{\bar{\beta}x^2}\right)\left(u_{H0}+\frac{\bar{u}_H}{1+x}\right)+(1+\underline{\beta}x)\left(1-\frac{l_1}{\bar{\beta}x}\right)\frac{-\bar{u}_H}{(1+x)^2}\right] \tag{17}$$

$$\begin{aligned}\frac{\partial v_2}{\partial x} =& \frac{\mu_1}{1-\frac{\underline{\beta}^2}{\bar{\beta}^2}}[\left(\frac{\underline{\beta}^2}{\bar{\beta}}+\frac{l_1}{\bar{\beta}x^2}\right)\left(u_{H0}+\frac{\bar{u}_L}{1+x}+\frac{\bar{u}_H-\bar{u}_L}{1+\frac{\underline{\beta}}{\bar{\beta}}x}\right)+ \\ & (1+\frac{\underline{\beta}^2}{\bar{\beta}}x)\left(1-\frac{l_1}{\bar{\beta}x}\right)\left(-\frac{\bar{u}_L}{(1+x)^2}-\frac{\bar{u}_H}{(1+\frac{\underline{\beta}}{\bar{\beta}}x)^2}\frac{\underline{\beta}}{\bar{\beta}}\right)]\end{aligned} \tag{18}$$

Step 1: show that the value functions v_1^* and v_2^* are continuously differentiable and the Envelope theorem applies.

The Lagrangian of the single-version problem (6) is

$$\mathscr{L}_1 = v_1(x,\bar{\beta},\underline{\beta}) - \gamma_1^1(\frac{l_1}{\bar{\beta}}-x) - \gamma_1^2(x-\frac{l_1}{\underline{\beta}}) \tag{19}$$

where γ's are the Lagrangian multipliers. By checking the first-order derivative with respect to x, it is easy to show the constraint $x \geq \frac{l_1}{\bar{\beta}}$ never binds. Hence, $\gamma_1^1 = 0$. Consider the Hessian matrix:

$$D^2\mathscr{L}_{x,\gamma_1^2} = \begin{vmatrix} \frac{\partial^2 v}{\partial x^2} & \frac{\partial^2 v}{\partial \gamma_1^2 \partial x} \\ \frac{\partial^2 v}{\partial x \partial \gamma_1^2} & \frac{\partial^2 v}{\partial (\gamma_1^2)^2} \end{vmatrix} = \begin{vmatrix} \frac{\partial^2 v}{\partial x^2} & -1 \\ -1 & 0 \end{vmatrix} = 1 > 0,$$

for at any $(x,\gamma_1^2) \in (0,\infty)^2$. According to Theorem 19.5 and Theorem 19.9 of Simon and Blume (1994), the value function v_1^* is continuously differentiable in $(\bar{\beta},\underline{\beta})$. Moreover, $\frac{\partial v_1^*}{\partial \bar{\beta}} = \frac{\partial v}{\partial \bar{\beta}}(x_1^*,x_1^*;\bar{\beta},\underline{\beta})$ and $\frac{\partial v_1^*}{\partial \underline{\beta}} = \frac{\partial v}{\partial \underline{\beta}}(x_1^*,x_1^*;\bar{\beta},\underline{\beta}) - \gamma_1^1(\bar{\beta},\underline{\beta})\frac{l_1}{\underline{\beta}^2}$.

Let the Lagrangian of the single-version problem (7) be

$$\mathscr{L}_2 = v(x, \frac{\bar{\beta}}{\underline{\beta}}x; \bar{\beta}, \underline{\beta}) - \gamma_2^1(\frac{l_1}{\bar{\beta}} - x) - \gamma_2^2(x - \frac{l_1\bar{\beta}}{\underline{\beta}^2}) \tag{20}$$

where γ's are the Lagrangian multipliers. With a similar argument, we can show the value function v_2^* is continuously differentiable in $(\bar{\beta}, \underline{\beta})$.

The constraint $x = \frac{l_1\bar{\beta}}{\underline{\beta}^2}$ will not bind whenever $v_2^* = v_1^*$ because $v_1(\frac{l_1}{\underline{\beta}}) > v_2(\frac{l_1\bar{\beta}}{\underline{\beta}^2})$. If the seller starts an immediate cascade, she will prefer a single-version strategy to avoid paying information rents.

Step 2: Noisy limit

Consider the single-version problem (6) first. $\forall x > 0$, we have $\frac{1-\frac{\underline{\beta}}{\bar{\beta}}}{\mu_1}\frac{\partial v_1}{\partial x} \xrightarrow[\underline{\beta}\to 1-0]{\bar{\beta}\to 1+0} \left(1 + \frac{l_1}{x^2}\right) u_{H0} + \frac{\bar{u}_H}{1+x}\left[\frac{l_1}{x^2} + \frac{l_1}{x}\right] > 0$. Hence, $\forall x > 0$ and $\boldsymbol{\beta}$ close enough to 1, $\frac{\partial v_1}{\partial x}(x, \boldsymbol{\beta}) > 0$. It follows that $x_1^* = \frac{l_1}{\underline{\beta}}$ around the noisy limit, which gives a profit $v_1^* = u_{H0} + \frac{\bar{u}_H}{1+l_1/\underline{\beta}}$.

Similarly, we can show that $\forall x > 0$, $\frac{1-\frac{\underline{\beta}^2}{\bar{\beta}^2}}{\mu_1}\frac{\partial v_2}{\partial x} \xrightarrow[\underline{\beta}\to 1-0]{\bar{\beta}\to 1+0} \left(1 + \frac{l_1}{x^2}\right) u_{H0} + \left(\frac{l_1}{x^2} + \frac{l_1}{x}\right)\frac{\bar{u}_H}{1+x} > 0$. Hence, $\forall x > 0$ and $\boldsymbol{\beta}$ close enough to $(1, 1)$, $\frac{\partial v_2}{\partial x}(x, \boldsymbol{\beta}) > 0$. $x_2^* = \frac{l_1\bar{\beta}}{\underline{\beta}^2}$ around the noisy limit, giving a profit $v_2^* = u_{H0} + \frac{\bar{u}_L}{1+\frac{l_1\bar{\beta}}{\underline{\beta}^2}} + \frac{\bar{u}_H - \bar{u}_L}{1+\frac{l_1}{\underline{\beta}}} < v_1^* = u_{H0} + \frac{\bar{u}_H}{1+l_1/\underline{\beta}}$.

Step 3: Optimality of a single-version strategy

We prove the following statement, which is also used in the proof of Theorem 2:

Claim 2 (Binding Constraint in the Limit). $\forall\bar{\beta} \in (1, \infty)$, *the constraint* $x \leq \frac{l_1}{\underline{\beta}}$ *binds for the single-version problem (6) in the limit* $\underline{\beta} \to 0 + 0$ *if and only if* $l_1 u_{H0} \geq \bar{\beta}\bar{u}_H$. *Furthermore, if* $l_1 u_{H0} \geq \bar{\beta}\bar{u}_H$, *for any* $\underline{\beta} \in (0, 1)$, $x \leq \frac{l_1\bar{\beta}}{\underline{\beta}^2}$ *binds for the limit basic-out problem (7) and thus* $v_2^*(\boldsymbol{\beta}) < v_1^*(\boldsymbol{\beta})$.

Proof. For any $x > 0$, we have

$$\lim_{\underline{\beta}\to 0+0} \frac{\partial v_1}{\partial x}\left(x, \bar{\beta}, \underline{\beta}\right) = \frac{\mu_1}{x^2\bar{\beta}(1+x)^2}\left[\left(l_1 u_{H0} - \bar{u}_H \bar{\beta}\right)x^2 + 2l_1 u_{H1} x + l_1 u_{H1}\right].$$

If $l_1 u_{H0} - \bar{u}_H \bar{\beta} \geq 0$, $\lim_{\underline{\beta}\to 0+0} \frac{\partial v_1}{\partial x}\left(x, \bar{\beta}, \underline{\beta}\right) \geq 0, \forall x > 0$ when $\underline{\beta}$ is sufficiently close to 0. It follows that $x_1^*(\boldsymbol{\beta}) = \frac{l_1}{\underline{\beta}}$ for small enough $\underline{\beta}$. If instead $l_1 u_{H0} - \bar{u}_H \bar{\beta} < 0$, $\frac{\partial v_1}{\partial x}\left(\frac{l_1}{\underline{\beta}}, \bar{\beta}, \underline{\beta}\right) = \frac{\mu_1}{\left(\frac{l_1}{\underline{\beta}}\right)^2\left(1+\frac{l_1}{\underline{\beta}}\right)^2 \bar{\beta}}\left[\left(l_1 u_{H0} - \bar{u}_H \bar{\beta}\right) l_1^2 + 2l_1^2 \underline{\beta} u_{H1} x + l_1 u_{H1} \underline{\beta}\right] < 0$ for small enough $\underline{\beta}$. $x = \frac{l_1}{\underline{\beta}}$ cannot be optimal.

From now on, assume $l_1 u_{H0} - \bar{u}_H \bar{\beta} \geq 0$. With a similar argument, we can show that the constraint $x \leq \frac{l_1 \bar{\beta}}{\underline{\beta}^2}$ also binds for $\underline{\beta}$ sufficiently close to 0.[18]

Next, we prove the following statement: if $\gamma_2^2(\bar{\beta}, \underline{\beta}) > 0$ for $\underline{\beta}$ sufficiently close to zero, $\forall \underline{\beta} \in (0,1), \gamma_2^2(\bar{\beta}, \underline{\beta}) > 0$. We will first show that the level curve at a constrained optimizer $\boldsymbol{x}^*(\bar{\beta}, \underline{\beta}) = (\frac{l_1 \bar{\beta}}{\underline{\beta}^2}, \frac{l_1}{\underline{\beta}})$ has a slope $\frac{dx_L}{dx_\Delta} > \frac{\bar{\beta}}{\underline{\beta}}$. Second, $\frac{dx_L}{dx_\Delta}$ increases in $\underline{\beta}$ whenever the slope gets sufficiently close to $\frac{\bar{\beta}}{\underline{\beta}}$. The final step will prove that the constraint $x \leq \frac{l_1 \bar{\beta}}{\underline{\beta}^2}$ must be binding for all possible $\underline{\beta} \in (0,1)$.

It is without loss to take $\ln v$ as our objective function because $v > 0$ except at $x_L = \frac{l_1}{\bar{\beta}}$ which is never optimal. The level curves are characterized by

$$\begin{aligned}\ln v\left(x_L, x_\Delta\right) = &\ln \mu_1 + \ln\left(1 + \underline{\beta} x_L\right) + \ln\left(\bar{\beta} x_L - l_1\right) - \ln\left(\bar{\beta} x_L - \underline{\beta} x_\Delta\right)\\ &+ \ln\left(u_{H0} + \frac{\bar{u}_L}{1+x_L} + \frac{\bar{u}_H - \bar{u}_L}{1+x_\Delta}\right) = C \text{ where } C \text{ can be any constant.}\end{aligned}$$

Notice that $\forall x_L > 0, x_\Delta = l_1/\underline{\beta}$, $\frac{\partial \ln v}{\partial x_L}(x_L, x_\Delta) = \frac{-\bar{u}_L}{(1+x_L)^2} / \left(u_{H0} + \frac{\bar{u}_L}{1+x_L} + \frac{\bar{u}_H - \bar{u}_L}{1+x_\Delta}\right) < 0$. It has two implications.

First, by the Implicit Function Theorem and v being continuously differentiable within the constraint set, there exists an open ball around the

[18] It requires a weaker condition, $l_1 u_{H0} - \bar{u}_L \bar{\beta} \geq 0$.

constrained optimizer where $\frac{\partial \ln v}{\partial x_L}(x_L, x_\Delta, \bar{\beta}, \underline{\beta}) < 0$ and the level curve can be thought of as the graph of a continuous function $x_L(x_\Delta)$. Its slope $g(x_L, x_\Delta, \bar{\beta}, \underline{\beta}) := \frac{dx_L}{dx_\Delta}(x_L, x_\Delta, \bar{\beta}, \underline{\beta})$ is defined by $\frac{\partial \ln v}{\partial x_L}\frac{dx_L}{dx_\Delta} + \frac{\partial \ln v}{\partial x_\Delta} = 0$.

On the horizontal line $x_\Delta = l_1/\underline{\beta}$, all the points left (right) to the constrained optimizer give a profit higher (lower) than at the optimizer. Also, $\gamma_2^2(\bar{\beta}, \underline{\beta}) > 0$ implies that along the constraint $x_\Delta = \frac{\underline{\beta}}{\bar{\beta}}x_L$, lower (higher) points to the constrained optimizer give lower (higher) profits than at the optimizer. Given that $\frac{\partial \ln v}{\partial x_\Delta}$ crosses zero at most once from below for all $\boldsymbol{x} \in (\frac{l_1}{\underline{\beta}}, \infty)^2$ (see the proof for Theorem 1), we must have $\frac{dx_L}{dx_\Delta}(\frac{l_1\bar{\beta}}{\underline{\beta}^2}, \frac{l_1}{\underline{\beta}}, \bar{\beta}, \underline{\beta}) > \frac{\bar{\beta}}{\underline{\beta}}$.

Using the Implicit Function Theorem again, we have

$$\frac{\partial g}{\partial \underline{\beta}}(x_L, x_\Delta, \bar{\beta}, \underline{\beta}) = -\frac{\frac{1}{\left(1+\underline{\beta}x_\Delta\right)^2} + \frac{\bar{\beta}x_L - \bar{\beta}x_\Delta g}{\left(\bar{\beta}x_L - \underline{\beta}x_0\right)^2}}{\frac{\partial \ln v}{\partial x_L}}$$

$$= -\frac{\frac{1}{\left(1+\underline{\beta}x_\Delta\right)^2} + \frac{\bar{\beta}x_L - \bar{\beta}x_\Delta g}{\left(\bar{\beta}x_L - \underline{\beta}x_\Delta\right)^2}}{\frac{\bar{\beta}}{\bar{\beta}x_L - l_1} - \frac{\bar{\beta}}{\bar{\beta}x_L - \underline{\beta}x_\Delta} + \frac{-\bar{u}_L}{(1+x_L)^2}\Big/\left(u_{H0} + \frac{\bar{u}_L}{1+x_L} + \frac{\bar{u}_H - \bar{u}_L}{1+x_\Delta}\right)}$$

It is continuous in $\boldsymbol{x}$ and $\underline{\beta}$ whenever $\frac{\partial \ln v}{\partial x_L} \neq 0$ and $\boldsymbol{x} \in \mathscr{C}(\boldsymbol{\beta})\backslash\{x_L = l_1/\bar{\beta}\}$. Moreover, when the slope approaches $\frac{\bar{\beta}}{\underline{\beta}}$ from above, $g \to \frac{\bar{\beta}}{\underline{\beta}} + 0$, and $x_\Delta = \frac{\underline{\beta}}{\bar{\beta}}x_L$, the numerator of $\frac{\partial g}{\partial \underline{\beta}}$ is positive as $-\bar{\beta}x_L + \bar{\beta}x_\Delta g = x_L(\bar{\beta} - \underline{\beta}g) \to 0$. So $\frac{\partial g}{\partial \underline{\beta}}(\frac{l_1\bar{\beta}}{\underline{\beta}^2}, \frac{l_1}{\underline{\beta}}, \bar{\beta}, \underline{\beta}) > 0$ at the constrained optimizer whenever $g(\frac{l_1\bar{\beta}}{\underline{\beta}^2}, \frac{l_1}{\underline{\beta}}, \bar{\beta}, \underline{\beta})$ is sufficiently close to $\frac{\bar{\beta}}{\underline{\beta}}$.

Suppose by contradiction that there exists a $\underline{\beta}' := \inf\{\underline{\beta} \in (0,1) : \gamma_2^2(\bar{\beta}, \underline{\beta}) = 0\}$. As implied by the Envelope theorem (See Step 1 of this section), the value function $v_2^*(\boldsymbol{\beta})$ and $x_2^*(\boldsymbol{\beta})$ are continuous in $\underline{\beta}$. Given that the constraint is binding for small enough $\underline{\beta}$, we can find a monotonically increasing sequence $\{g_n\}_{n=1}^{\infty}$, where $g_n := g(\frac{l_1\bar{\beta}}{\underline{\beta}_n^2}, \frac{l_1}{\underline{\beta}_n}, \bar{\beta}, \underline{\beta}_n), \forall n \geq 1$, such that $\lim_{n\to\infty} g_n = g' := \frac{\bar{\beta}}{\underline{\beta}'}$. This contradicts with the fact that $\frac{\partial g}{\partial \underline{\beta}}(\frac{l_1\bar{\beta}}{\underline{\beta}^2}, \frac{l_1}{\underline{\beta}}, \bar{\beta}, \underline{\beta}) > 0$ whenever $g(\frac{l_1\bar{\beta}}{\underline{\beta}^2}, \frac{l_1}{\underline{\beta}}, \bar{\beta}, \underline{\beta})$ is sufficiently close to $\frac{\bar{\beta}}{\underline{\beta}}$. Thus,

$\gamma_2^2(\bar{\beta}, \underline{\beta}) > 0$ for all possible $\underline{\beta} \in (0, 1)$.

□

To derive the single-version condition (8), note that $l_1 u_{H0} \geq \bar{\beta}\bar{u}_H \Leftrightarrow u_{H0} \geq \frac{\bar{\beta}}{1-\mu_1+\bar{\beta}}\mu_1 u_{H1}$. Since $\frac{\bar{\beta}}{1-\mu_1+\bar{\beta}} \in [\frac{1}{2-\mu_1}, 1]$ for all possible $\bar{\beta} \in [0, \infty]$, the single-version condition $u_{H0} \geq \mu_1 u_{H1}$ guarantees the optimality of a single-version strategy for all possible $\boldsymbol{\beta} \in (1, \infty) \times (0, 1)$.

Step 4: Informative limit

One implication of the previous step is that a single-version strategy is optimal around the informative limit under the single-version condition, $u_{H0} \geq \mu_1 u_{H1}$. Now we will show that when the condition fails, $v_2^*(\boldsymbol{\beta}) < v_1^*(\boldsymbol{\beta})$ as $\boldsymbol{\beta}$ gets sufficiently close to $(\infty, 0)$. Specifically, the main step is to prove that, if $\mu_1 u_{H1} > u_{H0} + \bar{u}_H/(1 + l_1/\underline{\beta})$, $\lim_{\bar{\beta}\to\infty}(v_2^*(\boldsymbol{\beta}) - v_1^*(\boldsymbol{\beta})) = 0$ and $\lim_{\bar{\beta}\to\infty}(\frac{\partial v_2^*}{\partial \bar{\beta}}(\boldsymbol{\beta}) - \frac{\partial v_1^*}{\partial \bar{\beta}}(\boldsymbol{\beta})) < 0$. As a result, given any $\underline{\beta} \in (0, 1)$, $v_2^*(\bar{\beta}, \underline{\beta}) - v_1^*(\bar{\beta}, \underline{\beta})$ approaches zero from above as $\bar{\beta}$ approaches infinity. It is then easy to verify that $u_{H0} < \mu_1 u_{H1} \Rightarrow \mu_1 u_{H1} > u_{H0} + \bar{u}_H/(1 + l_1/\underline{\beta})$ for small enough $\underline{\beta} > 0$.

Suppose $\mu_1 u_{H1} > u_{H0} + \bar{u}_H/(1 + l_1/\underline{\beta})$. Consider the single-version problem (6) first. According to (18), the first-order derivative is positive at $x = \frac{l_1}{\bar{\beta}}$, which implies $x_1^* > \frac{l_1}{\bar{\beta}}$.

Suppose $x_1^*(\boldsymbol{\beta})$ is bounded away from zero as $\bar{\beta} \to \infty$. $\frac{\partial v_1}{\partial x}$ approaches $\frac{\mu_1 \underline{\beta} u_{H0}}{(1+x)^2}[(1+x)^2 - \frac{\bar{u}_H(1-\underline{\beta})}{u_{H0}\underline{\beta}}]$ at any $x > 0$. The optimal x_1^* either gets arbitrarily close to the left boundary $\frac{l_1}{\bar{\beta}}$ which approaches zero in the limit and contradicts our assumption. Or the optimizer hits the right boundary $\frac{l_1}{\underline{\beta}}$ and generates a profit $v_1^* = u_{H0} + \bar{u}_H/(1 + l_1/\underline{\beta})$.

Suppose $x_1^*(\boldsymbol{\beta}) \to 0 + 0$ in the limit. Fix any $x > 0$ and we have $\lim_{\bar{\beta}\to\infty} \frac{\partial v_1}{\partial x} = \mu_1[(\underline{\beta} + \frac{l_1}{x^2\bar{\beta}}u_{H1})(u_{H0} + \frac{\bar{u}_H}{1+x}) + (1 + \underline{\beta}x)(1 - \frac{l_1}{\bar{\beta}x})\frac{-\bar{u}_H}{(1+x)^2}]$. The first observation is that $\bar{\beta}x_1^*$ cannot be bounded. Otherwise, the first-order derivative goes to (positive) infinity at $x \to 0 + 0$ which contradicts our assumption that $x_1^*(\boldsymbol{\beta}) \to 0+0$. So $\bar{\beta}x_1^* \to \infty$ and thus $v_1^* = \mu_1 u_{H1}$ in the limit. According to our assumption, this profit is higher than that obtained at $x = \frac{l_1}{\underline{\beta}}$. It follows that $\lim_{\bar{\beta}\to\infty} v_1^*(\bar{\beta}, \underline{\beta}) = \mu_1 u_{H1}$.

Similarly, we can show that $x_2^* \to 0+0$ and $v_2^* \to \mu_1 u_{H1}$ as $\bar{\beta} \to \infty$. To summarize, we have $\lim_{\bar{\beta}\to\infty}(v_2^*(\boldsymbol{\beta}) - v_1^*(\boldsymbol{\beta})) = 0$ and none of the constraints will bind as $\bar{\beta} \to \infty$. By the Envelope theorem, we can write

$$\frac{\partial v_2^*}{\partial \bar{\beta}} - \frac{\partial v_1^*}{\partial \bar{\beta}} = \frac{\mu_1}{\bar{\beta}} \left[\left(\frac{2 + \frac{l_1}{x_2^* \bar{\beta}^3} + \frac{\underline{\beta}^2 x_2^*}{\bar{\beta}}}{\left(1 - \frac{\underline{\beta}}{\bar{\beta}}\right)\left(1 + \frac{\underline{\beta}}{\bar{\beta}}\right)} - \frac{2\left(1 + \frac{\underline{\beta}^2}{\bar{\beta}} x_2^*\right)\left(1 - \frac{l_1}{x_2^* \bar{\beta}}\right)}{\left(1 - \frac{\underline{\beta}}{\bar{\beta}}\right)^2} \right) \left(u_{H0} + \frac{\bar{u}_L}{1 + x_2^*} + \frac{\bar{u}_H - \bar{u}_L}{1 + \frac{\underline{\beta}}{\bar{\beta}} x_2^*} \right) \right.$$
$$\left. + \frac{\left(1 + \frac{\underline{\beta}^2}{\bar{\beta}} x_2^*\right)\left(1 - \frac{l_1}{x_2^* \bar{\beta}}\right)}{1 - \frac{\underline{\beta}^2}{\bar{\beta}^2}} \underline{\beta} \frac{-(\bar{u}_H - \bar{u}_L)}{\left(1 + \frac{\underline{\beta}}{\bar{\beta}} x_2^*\right)^2} - \left(u_{H_0} + \frac{\bar{u}_H}{1 + x_1^*} \right) \frac{l_1 - \underline{\beta} x_1^*}{\bar{\beta} x_1^* - \underline{\beta} x_1^*} \frac{1 + \underline{\beta} x_1^*}{1 - \frac{\underline{\beta}}{\bar{\beta}}} \right]$$

The expression inside the square bracket approaches $-\underline{\beta}(\bar{u}_H - \bar{u}_L) < 0$ as $\bar{\beta} \to \infty$. Hence, $\lim_{\bar{\beta}\to\infty}(\frac{\partial v_2^*}{\partial \bar{\beta}}(\boldsymbol{\beta}) - \frac{\partial v_1^*}{\partial \bar{\beta}}(\boldsymbol{\beta})) < 0$.

A.6 Proof of Theorem 2

Here is the formal statement that we will prove. If $v_2^*(\boldsymbol{\beta}) - v_1^*(\boldsymbol{\beta})$ ever switches the sign on $\boldsymbol{\beta} \in (1, \infty) \times (0, 1)$,

1. fixing any $\underline{\beta} \in (0, 1)$, there exists a unique threshold $\bar{\beta}_0 \in (1, +\infty)$ such that $v_1^*(\bar{\beta}) \leq v_2^*(\bar{\beta})$ if and only if $\bar{\beta} \in [\bar{\beta}_0, +\infty)$;

2. fixing any $\bar{\beta} \in (1, \infty)$, there exists a unique threshold $\underline{\beta}_0 \in (0, 1)$ such that $v_1^*(\underline{\beta}) \leq v_2^*(\underline{\beta})$ if and only if $\underline{\beta} \in (0, \underline{\beta}_0]$.

The first step proves that it suffices to compare the value functions in a modified basic-out problem and the single version problem. Then, Step 2 and 3 prove the comparative statics results for $\bar{\beta}$ and $\underline{\beta}$ correspondingly.

Step 1: introduce a modified basic-out problem.

Given the value functions v_i^* are continuously differentiable (see the previous proof A.5), it suffices to show that $v_1^*(\boldsymbol{\beta}_0) = v_2^*(\boldsymbol{\beta}_0)$ whenever

$$\frac{\partial v_1^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0) < \frac{\partial v_2^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0) \text{ and } \frac{\partial v_1^*}{\partial \underline{\beta}}(\boldsymbol{\beta}_0) > \frac{\partial v_2^*}{\partial \underline{\beta}}(\boldsymbol{\beta}_0),$$

which will ensure the thresholds $\bar{\beta}_0$ and $\underline{\beta}_0$ are unique. Let $\boldsymbol{x}^0 := (x_1^0, x_2^0) := (x_1^*(\boldsymbol{\beta}_0), x_2^*(\boldsymbol{\beta}_0))$.

Now we introduce a modified seller's problem where we shut down the indirect effect from $\boldsymbol{\beta}$ on the optimal profits through the basic-out constraint ($x_\Delta = \frac{\underline{\beta}}{\bar{\beta}} x_L$).

$$\max_{x_L, x_\Delta \geq 0} v(x_L, x_\Delta) := \mu_1(1 + \underline{\beta} x_\Delta) \frac{\bar{\beta} x_L - l_1}{\bar{\beta} x_L - \underline{\beta} x_\Delta} [u_{H0} + \frac{\bar{u}_L}{1 + x_L} + \frac{\bar{u}_H - \bar{u}_L}{1 + x_\Delta}] \tag{21}$$

$$\text{s.t. } x_\Delta \leq x_L, \bar{\beta}_0 x_\Delta \geq \underline{\beta}_0 x_L \text{ and } \underline{\beta} x_\Delta \leq l_1 \leq \bar{\beta} x_L$$

Let $\hat{v}_2(x, \boldsymbol{\beta}, \boldsymbol{\beta}_0) := v(x, \frac{\underline{\beta}_0}{\bar{\beta}_0} x; \boldsymbol{\beta})$. The third argument $\boldsymbol{\beta}_0$ denotes the fixed β_0's in the basic-out constraint. The value function in the modified limit basic-out problem is then defined as

$$\hat{v}_2^*(\boldsymbol{\beta}, \boldsymbol{\beta}_0) := \max \hat{v}_2(x, \boldsymbol{\beta}, \boldsymbol{\beta}_0) \text{ s.t. } \frac{l_1}{\bar{\beta}} \leq x \leq \frac{l_1 \bar{\beta}_0}{\underline{\beta} \underline{\beta}_0}$$

and let $\hat{x}_2^*(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$ be the maximizer.

Since $v_1^*(\boldsymbol{\beta}_0) = v_2^*(\boldsymbol{\beta}_0) = \hat{v}_2^*(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0)$, we have (a) $\hat{x}_2^*(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) = x_2^*(\boldsymbol{\beta}_0) = x_2^0 \in (\frac{l_1}{\bar{\beta}}, \frac{l_1 \bar{\beta}_0}{\underline{\beta}_0^2})$ and (b) $\frac{\partial v}{\partial x_\Delta}(x_2^0, \frac{\underline{\beta}_0}{\bar{\beta}_0} x_2^0, \boldsymbol{\beta}_0) \leq 0$. As a result, $\frac{\partial v_2^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0) = \left[\frac{\partial}{\partial \bar{\beta}} v(x, \frac{\underline{\beta}}{\bar{\beta}} x; \boldsymbol{\beta})\right]|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, x=x_2^0} = \left[\frac{\partial v}{\partial x_\Delta}(x, \frac{\underline{\beta}}{\bar{\beta}} x; \boldsymbol{\beta}) \cdot (-\frac{\underline{\beta} x}{\bar{\beta}^2}) + \frac{\partial v}{\partial \bar{\beta}}(x, \frac{\underline{\beta}}{\bar{\beta}} x; \boldsymbol{\beta})\right]|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, x=x_2^0}$. $\frac{\partial \hat{v}_2^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0) = \left[\frac{\partial}{\partial \bar{\beta}} v(x, \frac{\underline{\beta}_0}{\bar{\beta}_0} x; \boldsymbol{\beta})\right]|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, x=x_2^0} = \frac{\partial v}{\partial \bar{\beta}}(x, \frac{\underline{\beta}_0}{\bar{\beta}_0} x; \boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, x=x_2^0} \leq \frac{\partial v_2^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0)$. Similarly, we can show that $\frac{\partial \hat{v}_2^*}{\partial \underline{\beta}}(\boldsymbol{\beta}_0) = \frac{\partial v}{\partial \underline{\beta}}(x, \frac{\underline{\beta}_0}{\bar{\beta}_0} x; \boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, x=x_2^0} \geq \frac{\partial v_2^*}{\partial \underline{\beta}}(\boldsymbol{\beta}_0) = \left[\frac{\partial v}{\partial x_\Delta}(x, \frac{\underline{\beta}}{\bar{\beta}} x; \boldsymbol{\beta}) \cdot (\frac{x}{\bar{\beta}}) + \frac{\partial v}{\partial \underline{\beta}}(x, \frac{\underline{\beta}_0}{\bar{\beta}_0} x; \boldsymbol{\beta})\right]|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0, x=x_2^0}$. Combining with the fact that the optimal profits are positive, it suffices to show

$$\frac{\partial \ln \hat{v}_2^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) > \frac{\partial \ln v_1^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0) \text{ and } \frac{\partial \ln \hat{v}_2^*}{\partial \underline{\beta}}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) < \frac{\partial \ln v_1^*}{\partial \underline{\beta}}(\boldsymbol{\beta}_0).$$

Moreover, by the Envelope theorem and $\hat{v}_2^*(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) = v_1^*(\boldsymbol{\beta}_0)$, we have

$$\frac{\partial \ln \hat{v}_2^*}{\partial \beta}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0) = \frac{\frac{\partial \hat{v}_2}{\partial \beta}(x_2^0, \boldsymbol{\beta}_0, \boldsymbol{\beta}_0)}{\hat{v}_2(x_2^0, \boldsymbol{\beta}_0, \boldsymbol{\beta}_0)} = \frac{\partial \ln \hat{v}_2}{\partial \beta}(x_2^0, \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) \text{ for all } \beta \in \{\bar{\beta}, \underline{\beta}\}$$

and

$$\frac{\partial \ln v_1^*}{\partial \bar{\beta}}(\boldsymbol{\beta}_0) = \frac{\partial \ln v_1}{\partial \bar{\beta}}(x_1^0, \boldsymbol{\beta}_0) \text{ and } \frac{\partial \ln v_1^*}{\partial \underline{\beta}}(\boldsymbol{\beta}_0) = \frac{\partial \ln v_1}{\partial \underline{\beta}}(x_1^0, \boldsymbol{\beta}_0) - \tilde{\gamma}_1^2(\boldsymbol{\beta}_0)\frac{l_1}{\underline{\beta}_0^2}$$

where $\tilde{\gamma}_1^2(\boldsymbol{\beta}) := \frac{\gamma_1^2(\boldsymbol{\beta})}{v_1(x_1^0, \boldsymbol{\beta})}$.

Step 2: $\frac{\partial \ln \hat{v}_2}{\partial \bar{\beta}}(x_2^0, \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) > \frac{\partial \ln v_1}{\partial \bar{\beta}}(x_1^0, \boldsymbol{\beta}_0)$**.**

Define $K(\boldsymbol{x}^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = \ln \hat{v}_2(x_2^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) - \ln v_1(x_1^0, \boldsymbol{\beta})$. Then $\frac{\partial K}{\partial \bar{\beta}}(\boldsymbol{x}^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = \frac{\partial \ln \hat{v}_2}{\partial \bar{\beta}}(x_2^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) - \frac{\partial \ln v_1}{\partial \bar{\beta}}(x_1^0, \boldsymbol{\beta}) = \frac{x_2^0}{\bar{\beta}x_2^0 - l_1} - \frac{1}{\bar{\beta} - \underline{\beta}\frac{\underline{\beta}_0}{\bar{\beta}_0}} - \frac{x_1^0}{\bar{\beta}x_1^0 - l_1} + \frac{1}{\bar{\beta} - \underline{\beta}}$.

Fixing any $\underline{\beta} \in (0,1)$, I want to show $\frac{\partial K}{\partial \bar{\beta}} \leq 0 \implies \frac{\partial^2 K}{\partial \bar{\beta}^2} > 0$.

$$\begin{aligned}
\frac{\partial^2 K}{\partial \bar{\beta}^2}(\boldsymbol{x}^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) &= \frac{\partial^2 \ln \hat{v}_2}{\partial \bar{\beta}^2}(x_2^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) - \frac{\partial^2 \ln v_1}{\partial \bar{\beta}^2}(x_1^0, \boldsymbol{\beta}) \\
&= -\left(\frac{x_2^0}{\bar{\beta}x_2^0 - l_1}\right)^2 + \frac{1}{(\bar{\beta} - \underline{\beta}\frac{\underline{\beta}_0}{\bar{\beta}_0})^2} + \left(\frac{x_1^0}{\bar{\beta}x_1^0 - l_1}\right)^2 - \frac{1}{(\bar{\beta} - \underline{\beta})^2} \\
&= \left(\frac{x_1^0}{\bar{\beta}x_1^0 - l_1} - \frac{x_2^0}{\bar{\beta}x_2^0 - l_1}\right)\left(\frac{x_1^0}{\bar{\beta}x_1^0 - l_1} + \frac{x_2^0}{\bar{\beta}x_2^0 - l_1}\right) - \\
&\left(\frac{1}{\bar{\beta} - \underline{\beta}} - \frac{1}{\bar{\beta} - \underline{\beta}\frac{\underline{\beta}_0}{\bar{\beta}_0}}\right)\left(\frac{1}{\bar{\beta} - \underline{\beta}} + \frac{1}{\bar{\beta} - \underline{\beta}\frac{\underline{\beta}_0}{\bar{\beta}_0}}\right) \\
&\geq \left(\frac{1}{\bar{\beta} - \underline{\beta}} - \frac{1}{\bar{\beta} - \underline{\beta}\frac{\underline{\beta}_0}{\bar{\beta}_0}}\right)\left(\frac{x_1^0}{\bar{\beta}x_1^0 - l_1} - \frac{1}{\bar{\beta} - \underline{\beta}} + \frac{x_2^0}{\bar{\beta}x_2^0 - l_1} - \frac{1}{\bar{\beta} - \underline{\beta}\frac{\underline{\beta}_0}{\bar{\beta}_0}}\right) \\
&> 0.
\end{aligned}$$

The first inequality follows from the fact that $\left(\frac{x_1^0}{\bar{\beta}x_1^0-l_1}+\frac{x_2^0}{\bar{\beta}x_2^0-l_1}\right)>0$ and $\frac{\partial K}{\partial \underline{\beta}}\leq 0$. The second inequality holds because $x_2^0<\frac{l_1\bar{\beta}_0}{\underline{\beta}\underline{\beta}_0}$ and $x_1^0\leq\frac{l_1}{\underline{\beta}}$.

Since $\frac{\partial K}{\partial \bar{\beta}}(\boldsymbol{x}^0,\boldsymbol{\beta},\boldsymbol{\beta}_0)$ is continuous in $\bar{\beta}$, it crosses zero at most once from below. Note also that $\lim_{\bar{\beta}\to\infty}\frac{\partial K}{\partial \bar{\beta}}(\boldsymbol{x}^0,\boldsymbol{\beta},\boldsymbol{\beta}_0)=0$. $\frac{\partial K}{\partial \bar{\beta}}(\boldsymbol{x}^0,\boldsymbol{\beta},\boldsymbol{\beta}_0)$ is thus either strictly positive or strictly negative on $\bar{\beta}\in(1,\infty)$.

Next, I want to prove

$$\lim_{\bar{\beta}\to\infty}K(\boldsymbol{x}^0,\bar{\beta},\underline{\beta}_0,\boldsymbol{\beta}_0)=\ln\left(\left(1+\frac{\underline{\beta}_0^2}{\bar{\beta}_0}x_2^0\right)\left(u_{H0}+\frac{\bar{u}_L}{1+x_2^0}+\frac{\bar{u}_H-\bar{u}_L}{1+\frac{\underline{\beta}_0}{\bar{\beta}_0}x_2^0}\right)\right)$$
$$-\ln\left((1+\underline{\beta}_0x_1^0)\left(u_{H0}+\frac{\bar{u}_H}{1+x_1^0}\right)\right)>0 \qquad (22)$$

, which, together with $K(\boldsymbol{x}^0,\bar{\beta}_0,\underline{\beta}_0,\boldsymbol{\beta}_0)=0$, implies $\frac{\partial K}{\partial \bar{\beta}}(\boldsymbol{x}^0,\bar{\beta},\underline{\beta}_0,\boldsymbol{\beta}_0)>0,\forall\bar{\beta}>1$. In particular, it holds true at $\bar{\beta}=\bar{\beta}_0$ and we complete Step 2.

Let $\lambda_1(x_1):=\frac{\bar{\beta}-\frac{l_1}{x_1}}{\bar{\beta}-\underline{\beta}}$ be the probability of a premium cascade conditional on state $\omega=1$. Likewise, define $\lambda_2(x_2):=\frac{\bar{\beta}-\frac{l_1}{x_2}}{\bar{\beta}-\frac{\underline{\beta}^2}{\bar{\beta}}}$. It is easy to verify that both functions are strictly increasing and have a range $[0,1]$. Hence, their inverse functions exist. We can then rewrite the profit functions for both problems as

$$v_i(\lambda)=v_i(\lambda_i^{-1}(\lambda))=\lambda_iP_i(\lambda),$$

for all $i\in\{1,2\}$, where $P_i(\lambda)$'s are pseudo price functions:

$$P_1(\lambda):=\mu_1\left(1+\underline{\beta}\frac{l_1}{\bar{\beta}-\lambda(\bar{\beta}-\underline{\beta})}\right)\left(u_{H0}+\frac{\bar{u}_H}{1+\frac{l_1}{\bar{\beta}-\lambda(\bar{\beta}-\underline{\beta})}}\right)$$
$$P_2(\lambda):=\mu_1\left(1+\frac{\underline{\beta}^2}{\bar{\beta}}\frac{l_1}{\bar{\beta}-\lambda(\bar{\beta}-\frac{\underline{\beta}^2}{\bar{\beta}})}\right)\left(u_{H0}+\frac{\bar{u}_L}{1+\frac{l_1}{\bar{\beta}-\lambda(\bar{\beta}-\frac{\underline{\beta}^2}{\bar{\beta}})}}+\frac{\bar{u}_H}{1+\frac{\underline{\beta}}{\bar{\beta}}\frac{l_1}{\bar{\beta}-\lambda(\bar{\beta}-\frac{\underline{\beta}^2}{\bar{\beta}})}}\right)$$

It turns out that the two pseudo-price functions cross each other at most once. I prove the following lemma in Appendix A.6.1.

Lemma 5. *There exists $\tilde{\lambda} \in [0,1]$ such that $P_2(\lambda) \leq P_1(\lambda)$ if and only if $\lambda \geq \tilde{\lambda}$.*

Let $\lambda_i^0 := \lambda_i(x_i^0)$ and $P_i^0 := P_i(\lambda_i^0), \forall i \in \{1,2\}$. The inequality (22) is then equivalent to $P_2^0 > P_1^0$, or equivalently $\lambda_2^0 < \lambda_1^0$.

First, we can show $\lambda_2^0 \leq \lambda_1^0$. Note that $\lambda_1^0 P_1^0 = \lambda_2^0 P_2^0 = \max_{\lambda \in [0,1]} \lambda P_1(\lambda) = \max_{\lambda \in [0,1]} \lambda P_2(\lambda)$ and $\lambda_2^0 < 1$. Since $P_2(\lambda_1^0)\lambda_1^0 \leq P_2^0 \lambda_2^0 = P_1^0 \lambda_1^0$, we must have $P_2(\lambda_1^0) \leq P_1^0 = P_1(\lambda_1^0)$. By Lemma 5, $P_2(\lambda) < P_1(\lambda), \forall \lambda > \lambda_1^0$. Thus, any λ higher than λ_1^0 cannot be λ_2^0. Formally, for any $\lambda > \lambda_1^0$, $\quad v_2(\lambda) = \lambda P_2(\lambda) < \lambda P_1(\lambda) = v_1(\lambda) \leq v_1(\lambda_1^0) = v_2(\lambda_2^0)$.

Suppose by contradiction that $\lambda_2^0 = \lambda_1^0$. Then, $P_2^0 = P_1^0$ and $P_2(\lambda)$ crosses $P_1(\lambda)$ from above at λ_1^0. It follows that $P_2'(\lambda_1^0) < P_1'(\lambda_1^0)$. But profit maximization requires $v_1'(\lambda_1^0) \leq 0 = v_2'(\lambda_2^0)$. Thus, $P_1'(\lambda_1^0)\lambda_1^0 + P_1(\lambda_1^0) \leq P_2'(\lambda_1^0)\lambda_1^0 + P_2(\lambda_1^0) \Rightarrow P_1'(\lambda_1^0) \leq P_2'(\lambda_1^0)$. Contradiction.

Step 3: $\frac{\partial \ln \hat{v}_2}{\partial \underline{\beta}}(x_2^0, \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) < \frac{\partial \ln v_1}{\partial \underline{\beta}}(x_1^0, \boldsymbol{\beta}_0) - \tilde{\gamma}_1^2(\boldsymbol{\beta}_0)\frac{l_1}{\underline{\beta}_0^2}$**.**

Define $\tilde{K}(\boldsymbol{x}^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = \ln \hat{v}_2(x_2^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) - \ln v_1(x_1^0, \boldsymbol{\beta}) - \tilde{\gamma}_1^2(\boldsymbol{\beta}_0)\frac{l_1}{\underline{\beta}}$. Then $\frac{\partial \tilde{K}}{\partial \underline{\beta}}(\boldsymbol{x}^0, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = \frac{\underline{\beta}_0 x_2^0/\bar{\beta}_0}{1+\underline{\beta} x_2^0 \frac{\underline{\beta}_0}{\bar{\beta}_0}} + \frac{\underline{\beta}_0/\bar{\beta}_0}{\bar{\beta}-\underline{\beta}\frac{\underline{\beta}_0}{\bar{\beta}_0}} - \frac{x_1^0}{1+\underline{\beta} x_1^0} - \frac{1}{\bar{\beta}-\underline{\beta}} + \tilde{\gamma}_1^2(\boldsymbol{\beta})\frac{l_1}{\underline{\beta}^2}$. As a first step, I want to show that, for any $\bar\beta \in (0,1)$, $\frac{\partial \tilde K}{\partial \underline\beta} \geq 0 \Rightarrow \frac{\partial^2 \tilde K}{\partial \underline\beta^2} < 0$. Suppose $\frac{\partial \tilde K}{\partial \underline\beta} \geq 0$.

$$\begin{aligned}
\frac{\partial^2 \tilde K^2}{\partial \underline\beta^2} = &\left(\frac{x_1^0}{1+\underline\beta x_1^0} - \frac{\underline\beta_0 x_2^0/\bar\beta_0}{1+\underline\beta x_2^0 \frac{\underline\beta_0}{\bar\beta_0}}\right)\left(\frac{x_1^0}{1+\underline\beta x_1^0} + \frac{\underline\beta_0 x_2^0/\bar\beta_0}{1+\underline\beta x_2^0 \frac{\underline\beta_0}{\bar\beta_0}}\right)\\
&- \left(\frac{1}{\bar\beta - \underline\beta} - \frac{\underline\beta_0/\bar\beta_0}{\bar\beta - \underline\beta\frac{\underline\beta_0}{\bar\beta_0}}\right)\left(\frac{1}{\bar\beta - \underline\beta} + \frac{\underline\beta_0/\bar\beta_0}{\bar\beta - \underline\beta\frac{\underline\beta_0}{\bar\beta_0}}\right) - \tilde\gamma_1^2(\boldsymbol\beta)\frac{2l_1}{\underline\beta^3}\\
\leq &- \left(\frac{1}{\bar\beta - \underline\beta} - \frac{\underline\beta_0/\bar\beta_0}{\bar\beta - \underline\beta\frac{\underline\beta_0}{\bar\beta_0}}\right)\left(\frac{x_1^0}{1+\underline\beta x_1^0} + \frac{\underline\beta_0 x_2^0/\bar\beta_0}{1+\underline\beta x_2^0 \frac{\underline\beta_0}{\bar\beta_0}}\right)\\
&- \left(\frac{1}{\bar\beta - \underline\beta} - \frac{\underline\beta_0/\bar\beta_0}{\bar\beta - \underline\beta\frac{\underline\beta_0}{\bar\beta_0}}\right)\left(\frac{1}{\bar\beta - \underline\beta} + \frac{\underline\beta_0/\bar\beta_0}{\bar\beta - \underline\beta\frac{\underline\beta_0}{\bar\beta_0}}\right)\\
&+ \tilde\gamma_1^2(\boldsymbol\beta)\frac{l_1}{\underline\beta^2}\left(\frac{x_1^0}{1+\underline\beta x_1^0} + \frac{\underline\beta_0 x_2^0/\bar\beta_0}{1+\underline\beta x_2^0 \frac{\underline\beta_0}{\bar\beta_0}} - \frac{2}{\underline\beta}\right) < 0, \forall \bar\beta \in (1,\infty), \underline\beta \in [0,1].
\end{aligned}$$

The first inequality follows from $\frac{\partial \tilde K}{\partial \underline\beta} \geq 0$. The second follows from (a) $\frac{1}{\bar\beta - \underline\beta} - \frac{\underline\beta_0/\bar\beta_0}{\bar\beta - \underline\beta\frac{\underline\beta_0}{\bar\beta_0}} > 0$; (b) $x_1^0 \leq l_1/\underline\beta$ and $x_2^0 \leq \frac{l_1\bar\beta_0}{\underline\beta\underline\beta_0}$; and (c) $\tilde\gamma_1^1(\boldsymbol\beta) \geq 0$.

Thus, for any given $\bar\beta \in (1,\infty)$, $\frac{\partial \tilde K}{\partial \underline\beta}(\boldsymbol x^0, \boldsymbol\beta, \boldsymbol\beta_0)$ crosses zero at most once and must be from above on $\underline\beta \in [0,1]$.

Given the assumption $v_1^*(\boldsymbol\beta_0) = v_1^*(\boldsymbol\beta_0)$ at some $\boldsymbol\beta_0 \in (1,\infty)\times(0,1)$, we know from Claim 2 that $l_1 u_{H0} < \bar u_H \bar\beta_0$. So $\lim_{\underline\beta\to 0+0} \tilde\gamma_1^2(\bar\beta_0, \underline\beta) = 0$ for small enough $\underline\beta$. The next step is to prove

$$\begin{aligned}
\lim_{\underline\beta\to 0+0} \tilde K(\boldsymbol x^0, \bar\beta_0, \underline\beta, \boldsymbol\beta_0) = \ln &\left(\mu_1\left(\frac{\bar\beta_0 x_2^0 - l_1}{\bar\beta_0 x_2^0}\right)\left(u_{H0} + \frac{\bar u_L}{1+x_2^0} + \frac{\bar u_H - \bar u_L}{1+\frac{\underline\beta_0}{\bar\beta_0}x_2^0}\right)\right)\\
&- \ln \mu_1\left(\left(\frac{\bar\beta_0 x_1^0 - l_1}{\bar\beta_0 x_1^0}\right)\left(u_{H0} + \frac{\bar u_H}{1+x_1^0}\right)\right) > 0 \quad (23)
\end{aligned}$$

Claim 3. $x_1^0 > \frac{\underline\beta_0}{\bar\beta_0} x_2^0$.

Proof. From the previous analysis we know $x_L = \frac{l_1}{\underline\beta}$ never binds and $v(x_L, x_\Delta) > 0, \forall x_L \neq \frac{l_1}{\underline\beta}$. For this proof, it is without loss to take $\ln v$

as our objective function in the general problem.

First, we prove that, given any feasible x_Δ, once $\frac{\partial \ln v}{\partial x_L}$ turns nonpositive (negative) at some $x_L > 0$, it stays nonpositive (negative) as x_L further increases. For $\boldsymbol{x} \in \mathscr{C}(\boldsymbol{\beta}) \backslash \{x_L = \frac{l_1}{\bar{\beta}}\}$, we have $\frac{\partial \ln v}{\partial x_L} \geq 0 \Leftrightarrow$

$$\begin{aligned}
&\left(\left(-\bar{\beta}\underline{\beta} x_\Delta + \bar{\beta} l_1\right) \left(u_{H0} + \frac{\bar{u}_H - \bar{u}_L}{1 + x_\Delta} \right) - \bar{u}_L \bar{\beta}^2 \right) x_L^2 \\
&+ \left[\left(-\bar{\beta}\underline{\beta} x_\Delta + \bar{\beta} l_1\right) \left[2 \left(u_{H0} + \frac{\bar{u}_H - \bar{u}_L}{1 + x_\Delta} \right) + \bar{u}_L \right] + \bar{u}_L \left(l_1 \bar{\beta} + \bar{\beta}\underline{\beta} x_\Delta \right) \right] x_L \\
&+ \left(-\bar{\beta}\underline{\beta} x_\Delta + \bar{\beta} l_1\right) \left[u_{H0} + \frac{\bar{u}_H - \bar{u}_L}{1 + x_\Delta} + \bar{u}_L \right] - \bar{u}_L l_1 \underline{\beta} x_\Delta \geq 0
\end{aligned}$$

The left-hand-side function is a quadratic function of x_L with a positive coefficient before x_L. It can be verified that the quadratic function is nonnegative at $x_L = \frac{l_1}{\bar{\beta}}$. Therefore, it is either nonnegative for all feasible x_L or strictly decreases from a nonnegative value. $\frac{\partial \ln v}{\partial x_L}(x_L, x_\Delta) \leq 0 \Rightarrow \frac{\partial \ln v}{\partial x_L}(x'_L, x_\Delta) < 0, \forall x'_L > x_L$.

Second, $\frac{\partial \ln v}{\partial x_L \partial x_\Delta} = \frac{-\bar{\beta}\underline{\beta}}{\left(\bar{\beta} x_L - \underline{\beta} x_0\right)^2} - \frac{\bar{u}_L}{(1+x_L)^2} \frac{(\bar{u}_H - \bar{u}_L)/(1+x_\Delta)^2}{\left(u_{H0} + \frac{\bar{u}_L}{1+x_L} + \frac{\bar{u}_H - \bar{u}_L}{1+x_\Delta}\right)^2} < 0, \forall \boldsymbol{x} \in \mathscr{C}(\boldsymbol{\beta}), 0 < \underline{\beta} < 1 < \bar{\beta}$.

For all $x_\Delta > x_1^0$ and $x_L \geq \frac{\bar{\beta}_0}{\underline{\beta}_0} x_1^0$, the optimality of x_0^1 implies $\frac{\partial \ln v}{\partial x_L}\left(x_1^0, x_1^0\right) \leq 0 \Rightarrow \frac{\partial \ln v}{\partial x_L}\left(\frac{\bar{\beta}_0}{\underline{\beta}_0} x_1^0, x_1^0\right) < 0 \Rightarrow \frac{\partial \ln v}{\partial x_L}\left(\frac{\bar{\beta}_0}{\underline{\beta}_0} x_1^0, x_\Delta\right) < 0 \Rightarrow \frac{\partial \ln v}{\partial x_L}\left(x_L, x_\Delta\right) < 0$. Any $x_2 \geq \frac{\bar{\beta}_0}{\underline{\beta}_0} x_1^0$ cannot be x_2^0 since an optimizer to the general problem (14) must satisfy $\frac{\partial \ln v}{\partial x_L}\left(x_2, \frac{\underline{\beta}_0}{\bar{\beta}_0} x_2\right) \geq 0$.

□

To complete the proof, use $v_1^*(\boldsymbol{\beta}_0) = v_1^*(\boldsymbol{\beta}_0)$ and Claim 3 and we have

$$\mu_1 \frac{\bar{\beta}_0 x_1^0 - l_1}{(\bar{\beta}_0 - \underline{\beta}_0) x_1^0} \left(u_{H0} + \frac{\bar{u}_H}{1 + x_1^0} \right) < \mu_1 \frac{\bar{\beta}_0 x_2^0 - l_1}{(\bar{\beta}_0 - \frac{\underline{\beta}_0^2}{\bar{\beta}_0}) x_2^0} \left(u_{H0} + \frac{\bar{u}_L}{1 + x_2^0} + \frac{\bar{u}_H - \bar{u}_L}{1 + \frac{\underline{\beta}_0}{\bar{\beta}_0} x_2^0} \right)$$

$$\Leftrightarrow \mu_1 \frac{\bar{\beta}_0 x_1^0 - l_1}{\bar{\beta}_0 x_1^0} \left(u_{H0} + \frac{\bar{u}_H}{1 + x_1^0} \right) < \mu_1 \frac{\bar{\beta}_0 x_2^0 - l_1}{\bar{\beta}_0 x_2^0} \frac{(\bar{\beta}_0 - \underline{\beta}_0)}{\bar{\beta}_0 - \frac{\underline{\beta}_0^2}{\bar{\beta}_0}} \left(u_{H0} + \frac{\bar{u}_L}{1 + x_2^0} + \frac{\bar{u}_H - \bar{u}_L}{1 + \frac{\underline{\beta}_0}{\bar{\beta}_0} x_2^0} \right)$$

$$< \mu_1 \frac{\bar{\beta}_0 x_2^0 - l_1}{\bar{\beta}_0 x_2^0} \left(u_{H0} + \frac{\bar{u}_L}{1 + x_2^0} + \frac{\bar{u}_H - \bar{u}_L}{1 + \frac{\underline{\beta}_0}{\bar{\beta}_0} x_2^0} \right).$$

A.6.1 Proof of Lemma 5

Let $A(\lambda) := \bar{\beta} - \lambda \left(\bar{\beta} - \frac{\underline{\beta}^2}{\bar{\beta}} \right)$ and $B(\lambda) := \bar{\beta} - \lambda(\bar{\beta} - \underline{\beta})$.

$$P_2(\lambda) \leq P_1(\lambda) \Leftrightarrow$$

$$\mu_1 \left(1 + \frac{\underline{\beta}^2}{\bar{\beta}} \frac{l_1}{\bar{\beta} - \lambda(\bar{\beta} - \frac{\underline{\beta}^2}{\bar{\beta}})} \right) \left(u_{H0} + \frac{\bar{u}_L}{1 + \frac{l_1}{\bar{\beta} - \lambda(\bar{\beta} - \frac{\underline{\beta}^2}{\bar{\beta}})}} + \frac{\bar{u}_H}{1 + \frac{\underline{\beta}}{\bar{\beta}} \frac{l_1}{\bar{\beta} - \lambda(\bar{\beta} - \frac{\underline{\beta}^2}{\bar{\beta}})}} \right)$$

$$- \mu_1 \left(1 + \underline{\beta} \frac{l_1}{\bar{\beta} - \lambda(\bar{\beta} - \underline{\beta})} \right) \left(u_{H0} + \frac{\bar{u}_H}{1 + \frac{l_1}{\bar{\beta} - \lambda(\bar{\beta} - \underline{\beta})}} \right) \leq 0$$

Rearrange the inequality and, using the fact that $B + l_1 > 0, l_1 > 0$, we have $P_2(\lambda) \leq P_1(\lambda) \Leftrightarrow$

$$\bar{u}_H l_1 (1 - \underline{\beta}) \leq$$

$$\bar{u}_L \left(1 - \frac{\underline{\beta}^2}{\bar{\beta}} \right) \frac{B + l_1}{A + l_1} + (\bar{u}_H - \bar{u}_L) \frac{\underline{\beta}}{\bar{\beta}} (1 - \underline{\beta}) \frac{B + l_1}{A + \frac{\underline{\beta} l_1}{\bar{\beta}}} + u_{H0} \underline{\beta} \frac{A - \frac{\underline{\beta}}{\bar{\beta}} B}{AB} (B + l_1)$$

Let's define $h_1(\lambda) := \bar{u}_H l_1 (1 - \underline{\beta}) - \bar{u}_L \left(1 - \frac{\underline{\beta}^2}{\bar{\beta}} \right) \frac{B + l_1}{A + l_1} - (\bar{u}_H - \bar{u}_L) \frac{\underline{\beta}}{\bar{\beta}} (1 - \underline{\beta}) \frac{B + l_1}{A + \frac{\underline{\beta} l_1}{\bar{\beta}}}$ and $h_2(\lambda) := u_{H0} \underline{\beta} \frac{A - \frac{\underline{\beta}}{\bar{\beta}} B}{AB} (B + l_1)$. We want to prove $h(\lambda) = h_1(\lambda) - h_2(\lambda)$ crosses zero on $\lambda \in [0, 1]$ at most once and must be from above.

The next claim collects several useful properties of functions $h_1, h_2, A,$ and B.

Claim 4. *For all $\lambda \in [0,1]$,*

1. $0 < -B'(\lambda) < -A'(\lambda)$ *and* $0 < A(\lambda) < B(\lambda)$.

2. $h_1'(\lambda) < 0$. $h_1''(\lambda) > \frac{-2A'}{A} h_1'(\lambda)$.

3. $h_2(\lambda) \geq 0$. *The equality holds only at* $\lambda = 1$. $h_2''(\lambda) < \frac{-2A'}{A} h_2'(\lambda)$.

Proof. I will prove $h_1''(\lambda) > \frac{-2A'}{A} h_1'(\lambda)$ and $h_2''(\lambda) < \frac{-2A'}{A} h_2'(\lambda)$. The other results are easy to verify.

$$\begin{aligned}
&h_1''(x) = -\bar{u}_L \left(1 - \frac{\underline{\beta}^2}{\bar{\beta}}\right) \left(\bar{\beta} + l_1\right) \underline{\beta} \left(1 - \frac{\underline{\beta}^2}{\bar{\beta}}\right) \frac{-2A'}{(A+l_1)^3} \\
&- (\bar{u}_H - \bar{u}_L) \frac{\underline{\beta}}{\bar{\beta}} (1 - \underline{\beta}) \left(l_1 + \underline{\beta}\right) (\bar{\beta} - \underline{\beta}) \frac{-2A'}{\left(A + \frac{\underline{\beta} l_1}{\bar{\beta}}\right)^3} \\
&= \frac{-2A'}{A + \frac{\underline{\beta} l_1}{\bar{\beta}}} [\underbrace{-\bar{u}_L \left(1 - \frac{\underline{\beta}^2}{\bar{\beta}}\right) \left(\bar{\beta} + l_1\right) \underline{\beta} \left(1 - \frac{\underline{\beta}^2}{\bar{\beta}}\right) \frac{1}{(A+l_1)^2}}_{<0} \frac{A + \frac{\underline{\beta} l_1}{\bar{\beta}}}{A + l_1} \\
&- (\bar{u}_H - \bar{u}_L) \frac{\underline{\beta}}{\bar{\beta}} (1 - \underline{\beta}) \left(l_1 + \underline{\beta}\right) (\bar{\beta} - \underline{\beta}) \frac{1}{\left(A + \frac{\underline{\beta} l_1}{\bar{\beta}}\right)^2}] \\
&\Rightarrow h_1''(\lambda) > \frac{-2A'}{A + \frac{\underline{\beta} l_1}{\bar{\beta}}} h_1'(\lambda) > \frac{-2A'}{A} h_1'(\lambda)
\end{aligned}$$

The first inequality follows from the formula for $h_1'(\lambda)$ and $0 < \frac{A + \frac{\underline{\beta} l_1}{\bar{\beta}}}{A + l_1} < 1$. The second inequality follows from $h_1' < 0$, $\frac{\underline{\beta} l_1}{\bar{\beta}} > 0$ and $A' < 0$.

Likewise, $-\frac{2A'}{A} h_2'(\lambda) - h_2''(\lambda) = u_{H0} \frac{-2A'}{A} \left[(B + l_1)(-\frac{B'}{B^2}) \left(1 - \frac{B'}{B}\frac{A}{A'}\right) + \frac{B'}{B} \left(1 - \frac{B'}{B}\frac{A}{A'}\right)\right] = u_{H0} \frac{-2A'}{A} \left(1 - \frac{B'}{B}\frac{A}{A'}\right) (-\frac{B'}{B}) \frac{l_1}{B} > 0.$

□

As a result, $h'(\lambda) = 0 \Rightarrow h''(\lambda) = h_1''(\lambda) - h_2''(\lambda) > \frac{-2A'}{A}(h_1'(\lambda) - h_2'(\lambda)) = 0$. $h(\lambda)$ is either monotone or decreases first and then increases on $[0,1]$.

Also, it is easy to derive that $h(1) = \bar{u}_L(1-\underline{\beta})\left(1 - \frac{1-\frac{\underline{\beta}^2}{\bar{\beta}}}{1-\underline{\beta}}\frac{\underline{\beta}+l_1}{\frac{\underline{\beta}^2}{\bar{\beta}}+l_1}\right) < 0$. If $h(0) \leq 0, h(\lambda) \leq 0, \forall \lambda \in [0,1]$. Otherwise, $h(0) > 0 > h(1), h(\lambda)$ crosses zero once from above on $[0,1]$.

References

Adams, W. J. and J. L. Yellen (1976). Commodity bundling and the burden of monopoly. *The Quarterly Journal of Economics 90*(3), 475–498. Publisher: Oxford University Press.

Anderson, E. T. and J. D. Dana (2009). When is price discrimination profitable? *Management Science 55*(6), 980–989.

Arcidiacono, P., P. B. Ellickson, C. F. Mela, and J. D. Singleton (2020). The competitive effects of entry: Evidence from supercenter expansion. *American Economic Journal: Applied Economics 12*(3), 175–206.

Arieli, I., M. Koren, and R. Smorodinsky (2022). The implications of pricing on social learning. *Theoretical Economics 17*(4), 1761–1802.

Bergemann, D. and J. Välimäki (1997). Market diffusion with two-sided learning. *The RAND Journal of Economics*, 773–795.

Bergemann, D. and J. Välimäki (2000). Experimentation in markets. *The Review of Economic Studies 67*(2), 213–234.

Bergemann, D. and J. Välimäki (2002). Entry and vertical differentiation. *Journal of Economic Theory 106*(1), 91–125.

Bergemann, D. and J. Välimäki (2006). Dynamic pricing of new experience goods. *Journal of Political Economy 114*(4), 713–743.

Bonatti, A. (2011). Menu pricing and learning. *American Economic Journal: Microeconomics 3*(3), 124–63.

Bose, S., G. Orosel, M. Ottaviani, and L. Vesterlund (2006). Dynamic monopoly pricing and herding. *The RAND Journal of Economics 37*(4), 910–928.

Bose, S., G. Orosel, M. Ottaviani, and L. Vesterlund (2008). Monopoly pricing in the binary herding model. *Economic Theory 37*(2), 203–241.

DellaVigna, S. and M. Gentzkow (2019). Uniform pricing in us retail chains. *The Quarterly Journal of Economics 134*(4), 2011–2084.

Fan, Z., X. Weng, L.-A. Zhou, and Y. Zhou (2023). Observational learning and information disclosure in search markets. *Working Paper*.

Ghili, S. (2023). A characterization for optimal bundling of products with nonadditive values. *American Economic Review: Insights 5*(3), 311–326.

Haghpanah, N. and J. Hartline (2021). When is pure bundling optimal? *The Review of Economic Studies 88*(3), 1127–1156.

Johnson, J. P. and D. P. Myatt (2006). On the Simple Economics of Advertising, Marketing, and Product Design. *The American Economic Review 96*(3).

Laiho, T. and J. Salmi (2021). Coasian dynamics and endogenous learning. *Working Paper*.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory 18*(2), 301–317.

Newberry, P. W. (2016). An empirical study of observational learning. *The RAND Journal of Economics 47*(2), 394–432.

Phillips, R. L. (2015). *Pricing and Revenue Optimization: Second Edition.* Stanford University Press.

Reimers, I. and J. Waldfogel (2017). Throwing the books at them: Amazon's puzzling long run pricing strategy. *Southern Economic Journal 83*(4), 869–885.

Salant, S. W. (1989). When is inducing self-selection suboptimal for a monopolist? *The Quarterly Journal of Economics 104*(2), 391–397.

Sandmann, C. (2023). When are single-contract menus profit maximizing. *Working Paper*.

Simon, C. P. and L. Blume (1994). *Mathematics for economists*, Volume 7. Norton New York.

Smith, L. and P. Sørensen (2000). Pathological outcomes of observational learning. *Econometrica 68*(2), 371–398.

Stigler, G. J. (1963). United States v. Loew's Inc.: A Note on Block-Booking. *The Supreme Court Review 1963*, 152–157. Publisher: University of Chicago Press.

Stokey, N. L. (1979). Intertemporal price discrimination. *The Quarterly Journal of Economics 93*(3), 355–371.

Welch, I. (1992). Sequential sales, learning, and cascades. *Journal of Finance 47*(2), 695–732.

Yang, F. (2024). Nested bundling. *Working Paper*.

Zhang, J. (2010). The sound of silence: Observational learning in the US kidney market. *Marketing Science 29*(2), 315–335.

Zhang, J. and P. Liu (2012). Rational herding in microloan markets. *Management Science 58*(5), 892–912.