Melons as Lemons: Asymmetric Information, Consumer Learning and Quality Provision*

Jie Bai (Harvard Kennedy School) †

July 2018

Abstract

There is often a lack of reliable quality provision in many markets in developing countries and firms generally lack a reputation for quality. One potential explanation is that mistrust due to past bad behavior can make reputation-building difficult. I examine this hypothesis in a setting that features typical market conditions in developing countries: the retail watermelon markets in a major Chinese city. I first demonstrate empirically that there is substantial asymmetric information between sellers and buyers on quality and a stark absence of quality premium at baseline. I then randomly introduce one of two branding technologies into 40 out of 60 markets–one sticker label that is widely used and counterfeited and one novel expensive laser-cut label. The experiment findings show that laser-branding induced sellers to provide higher quality and led to higher sales profits. However, after the intervention was withdrawn, all markets reverted back to baseline. I incorporate the experimental variation into an empirical model of consumer learning and seller reputation building. The results suggest that consumers are hesitant to upgrade their perception under stickers, which makes reputation-building a low-return investment. While the new technology enhances learning, the resulting increase in profits is not sufficient to cover the fixed cost of the technology for small individual sellers. Counterfactual analysis shows that information friction and fragmented market lead to significant under-provision of quality. (*JEL*: D22, D83, L11, L14, L15, O10, O12)

---

*The paper was previously circulated under the title "Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation." I am very grateful to Benjamin Olken, Abhijit Banerjee, Robert Townsend and Nikhil Agarwal for their continuous encouragement and guidance throughout this project. I thank David Atkin, Esther Duflo, Marcel Fafchamps, Hanming Fang, Glenn Ellison, Robert Gibbons, German Sergio Gieczewski, Seema Jayachandran, Yan Ji, Siyuan Liu, Rocco Macchiavello, Yuhei Miyauchi, Xiaosheng Mu, Ariel Pakes, Frank Schilbach, Peng Shi, Maheshwor Shrestha, Paulo Somaini, Tavneet Suri, Jean Tirole, Juuso Toikka, Richard Zeckhauser, Juanjuan Zhang, Hongkai Zhang, and seminar participants at NEUDC, AFE, Berkeley, Chicago (Economics/Booth), CICER, Columbia, Cornell, GWU, Harvard Economics, Harvard Kennedy School, Microsoft Research, MIT Sloan Applied Economics, MIT Sloan Marketing, Northwestern, NYU Stern, PKU, PSU, SITE, Stanford GSB, UMich, UPenn, Yale SOM Marketing, and the development and IO lunches at MIT for helpful feedback. Funding for this project was generously provided by the Weiss Family Fund, the Abdul Latif Jameel Poverty Action Lab, and the George and Obie Shultz Fund. All errors are my own.

†Address: 79 JFK Street, Cambridge, MA 02138. E-mail: jie_bai@hks.harvard.edu.

1 Introduction

A key problem in developing countries is the lack of reliable provision of high quality goods and services. The problem is exacerbated in markets of experience goods, such as food products and pharmaceuticals (WHO, 2007). In recent years, there have been growing concerns among the public regarding product quality in developing countries.[1] At the heart of the issue is an information problem: when contracting on quality is difficult, information frictions can lead to quality deterioration and firms need a good reputation to succeed. However, a reputation for quality is precisely what many firms in developing countries are lacking. The question is, then, what are the barriers that hinder firms' ability and incentive to establish reputation for quality?

There are many possible explanations, including the lack of technology to produce quality, the lack of quality inputs, poor access to credit, or poor management. In this paper, I focus on a potential demand-side constraint: the lack of trust. For example, a series of recent quality scandals over the past few years have led to enormous mistrust among consumers in China.[2] In a recent survey of over 600 Chinese manufacturing firms, "mistrust" was cited as one of the main challenges for penetrating higher-end markets.[3] In this paper, I argue that such pervasive generalized mistrust can make reputation building a difficult and low-return investment: because of the information problem, a firm's claim of offering high quality cannot be immediately verified, and therefore consumers' perception and speed of learning matter for the reputational incentive; firms rationally discounting future profits may lack the incentive to provide quality if trust takes a long time to build. In such environments, credible signaling technologies can play a crucial role. Various technologies have been tried in the past, ranging from the simplest way of charging higher prices to fancier packaging, to various certified labels. However, rampant counterfeiting has undermined these strategies and many markets have repeatedly reversed back to a low-quality equilibrium.[4] Over time, beliefs can become very pessimistic and self-confirming.

I first build a simple model to illustrate the interaction between consumer learning and seller

[1] In China, food safety and quality has been identified as one of the top 10 concerns of Chinese people in the 19th Party Congress (see http://news.xinhuanet.com/politics/19cpcnc/2017-10/21/c_1121836409.htm).

[2] For example, see an article by World Policy about the food scandals and consequences in China: http://www.worldpolicy.org/blog/2013/08/07/food-safety-and-china-scandal-and-consequence

[3] The survey is led by Jinan Institute for Economic and Social Research (IESR) and the Guangzhou General Administration of Quality Supervision, Inspection and Quarantine. I thank IESR for sharing the data.

[4] In the food sector, various certified labels including "pollution-free", "green" and "Organic" have been widely forged and the system has failed to gain consumers' trust (http://finance.huanqiu.com/pictures/2011-10/2127997.html).

trust building. To test the theory, I examine a setting where there was a lack of quality provision at baseline. I randomly introduced different signaling technologies and a monetary incentive that were strong enough to break the baseline equilibrium and induce quality provision. The experimental variation combined with rich data collected from both sides of the market allows me to diagnose the underlying reasons for the baseline lack of quality. The results show that information frictions lead to significant under-provision of quality in this setting.

The setting is the local food markets in a major Chinese city, an important institution familiar to many developing countries and the final link in the long supply chain for many agricultural products (AAFC, 2014). I focus on fruit stores selling watermelons, one of the most popular summer fruits. Common to many other food products, watermelon is a typical experience good. However, a key advantage is that a watermelon's quality can be well-captured by its sweetness, which can be measured (ex-post) using a sweet meter.[5] This allows me to *directly* examine sellers' incentive to provide quality and consumers' demand for quality. I first document substantial asymmetric information between sellers and buyers on this key dimension of quality and a stark absence of a quality-price premium at baseline.

The experiment involves 60 sellers in 60 different markets in Shijiazhuang, China. The large number of independent local markets allows randomization at the market level. I randomly introduced one of two branding technologies into 40 out of the 60 markets—one sticker label that is widely used and often counterfeited, and one novel expensive laser-cut label. Pilot surveys suggest that consumers regard laser-branding as being more effective at deterring counterfeits because laser machines are very expensive. Hence, the new technology could potentially dispel negative stereotypes associated with stickers, thereby allowing sellers to establish trust faster.[6] For a cross-randomized subset of sellers, I further provided a temporary monetary incentive to invest in high quality of their watermelons. The incentive treatment helps to shed light on the underlying reputational forces and also helps to identify the costs of providing quality, which is hard to measure directly. The intervention lasted over eight weeks, spanning the entire peak season for watermelons. I kept track of sellers' quality, pricing and sales over the entire season, and collected household panel purchasing data to examine the demand side's response.

There are three main experimental findings. First, laser branding induced sellers to pro-

[5] I conducted a baseline blind tasting test. The test shows that sweetness strongly correlates with consumer's taste: among 210 consumers who were asked to compare two watermelons of high and low sweetness measures, 97% preferred the sweeter one.

[6] This relates to the collective reputation theory in Tirole (1996) and information free-riding in Fang (2001).

vide a genuine quality-price premium, establishing that reputational incentives can potentially motivate quality. On the other hand, evidence for the sticker group is quite mixed: on average, quality of the premium pile was not significantly higher than the market average. Second, the incentive treatment successfully induced sellers to provide higher quality than their non-incentivized counterparts, but higher quality was only sustained for the laser incentive group. Third, in terms of sales outcomes, sticker did not outperform the baseline. In contrast, sellers in the laser group earned 30-40% higher sales profits on average as a result of both higher prices and higher total sales. This result demonstrates that there is a high demand for quality and reputation can be made to pay. Having said that, one year after the intervention when the laser technology was no longer provided for free, all markets reverted back to baseline. This suggests that small individual sellers would not have the incentive to invest in the new technology themselves. This is consistent with sellers' self-reported willingness to pay for different technologies elicited at the endline.

The experimental findings provide a qualitative explanation for the lack of quality provision at baseline. Additional evidence exploring the dynamics of household purchasing patterns and sales trajectories further point to the role of initial beliefs, learning and signaling in the presence of information problems. I discuss alternative models of relational contracting, a fixed laser-coolness effect, and a coordination mechanism (Klein and Leffler, 1981) using supplementary data collected from these markets, including sellers' peach sales and household endline perceptions. Overall, the findings support the model of learning and trust building.

I next incorporate the experimental variation into an empirical model to recover the underlying evolution beliefs under different branding technologies. The model is estimated using simulated maximum likelihood and I exploit purchasing patterns and experience realizations observed in the household panel data for identification. In line with the experimental findings, the structural estimates indicate that consumers' prior perception is more "stubborn" under sticker branding than under laser branding. As a result, trust can take a long time to establish, which explains why sellers do not have the incentive to provide quality without the intervention. To shed light on sellers' incentives under the new technology, I integrate the demand model with a supply model to recover the unobserved effort costs and discount factor from observed price and quality decisions using a minimum distance estimator. The estimates reveal that while the new technology enhances consumer learning and thereby strengthens sellers' incentives, the increase in the discounted return, taking into account effort costs, is still not large enough to justify the fixed cost of the technology for individual sellers. There are two reasons: (a) each seller's size is very small; and (b) it may be difficult for sellers to extract all the consumer surplus due to market competition.

To further highlight the interaction between information friction and market structure and tradeoffs faced by policy makers, I conduct several counterfactual exercises to examine the role of firm size and market competition. The results indicate that asymmetric information and fragmented market lead to significant under-provision of quality in this setting. While an individual seller would not undertake such costly investment, a third-party could invest in the new technology and subsidize it for sellers to improve aggregate welfare. Alternatively, since sellers' net profits scale up with market size, the results suggest that there could be a profitable entry opportunity for a large upstream firm.

While the specific takeaways are product-location specific as the exact learning dynamics and quality provision technologies vary across industries and settings, the key underlying features–asymmetric information, rising demand for quality, widespread mistrust, small firm size and fragmented market–are common to many product markets in developing countries. The lesson highlighted in this study could be more pertinent in settings where firms are more mobile and information arrives slowly. For many quality and safety issues, consumers do not learn unless something catastrophic happens. Thus, markets can easily get stuck in a low-trust-low-quality equilibrium since "good news" arrives very slowly, if at all; the welfare loss due to the information problem could be much larger.

This paper contributes to the empirical literature on consumer learning, firm reputation and quality provision in markets with information problems.[7] While many studies examine online trading environments,[8] empirical work in the offline world is relatively sparse (Banerjee and Duflo, 2000; Jin and Leslie, 2009; Macchiavello, 2010; List, 2006; Bardhan, Mookherjee, and Tsumagari, 2013; Björkman-Nyqvist, Svensson, and Yanagizawa-Drott, 2013; Macchiavello and Morjaria, 2015). As discussed in Bar-Isaac and Tadelis (2008), the empirical challenge is that researchers typically do not observe all information available to buyers, and sellers' behavior beyond what the buyers observe. This study takes advantage of a field experiment that directly keeps track of the both sides. The results demonstrate that the way consumers gather

[7]This study is motivated by the extensive body of economics and marketing literature on the role of advertising as signals for product quality (Bagwell, 2007). Most theoretical work focuses on equilibrium predictions between advertising and quality, where quality is exogenous.

[8]For example, see Jin and Kato (2006); Cabral and Hortacsu (2010); Klein, Lambertz, and Stahl (2016)

information and learn shapes seller's incentive. This recalls the finding in Björkman-Nyqvist, Svensson, and Yanagizawa-Drott (2013) that quality provision of anti-malaria drugs in Uganda is hampered by consumers' misconceptions. Although the contexts differ, the policy conclusions are alike. To motivate high quality provision, policies that enhance consumer learning or entry of large firms may be needed.

Findings of the study also speak to the role of firm size and market structure on product quality. Most empirical work focuses on settings where quality is observable (e.g., see Kugler and Verhoogen (2012)). This study examines a setting with information asymmetry. Jensen and Miller (2016) also studies informational barriers as a potential constraint on firm growth, focusing on firm size. The current study takes the small firm size and entry in each market as given, and examine how such features may interact with the information problem to hinder quality provision.[9] The modeling framework is related to the demand accumulation model in Foster, Haltiwanger, and Syverson (2016). I focus on consumer learning about quality in a vertically differentiated market as the fundamental underlying force of firm growth.

The study also relates to the broad literature on firm growth and quality upgrading in development and trade.[10] Previous studies have addressed: (1) supply side constraints, including credit access, lack of quality inputs, managerial constraints, and interfirm relationships;[11] and (2) demand side factors, including access to high-income markets (e.g., Verhoogen (2008); Atkin, Khandelwal, and Osman (2017)). This study highlights another potential barrier to quality upgrading, which is the information problem and mistrust.[12] Such mistrust, often targeted at a broad group level (industry or country), generates an important externality that not only hinders individual firm's ability to penetrate higher-end markets (as we see for sellers in the sticker group) but also hurts new firms which are "endowed" with the damaged reputation of the ancestors (Macchiavello, 2010).

The remainder of this paper is organized as follows. Section 2 describes the setting. Section 3 outlines the model. Section 4 describes the experimental design and the data. Section 5 presents the experimental results. Section 6 estimates an empirical model of learning and quality provision. Section 7 uses the structural estimates to examine the welfare implications of

[9]For theoretical insights, see Kranton (2003) and Villas-Boas (2004). Macchiavello, Morjaria, et al. (2016) also considers an asymmetric information setting and shows competition can weaken the value of relationships.

[10]See De Loecker and Goldberg (2014) for a comprehensive review of the empirical literature.

[11]E.g., De Mel, McKenzie, and Woodruff (2008); Harrison and Rodríguez-Clare (2009); Kugler and Verhoogen (2012); Banerjee (2013); Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013); Cai and Szeidl (2017).

[12]Studies have found that information frictions are important in trade–e.g., see Allen (2014); Startz (2016)

information friction and fragmented market. Section 8 concludes. An online appendix contains additional technical results pertaining to the theoretical analysis and empirical estimation.[13]

2 Setting

2.1 Local Retail Markets

Despite the rise of supermarkets and e-commerce, semi-formal, open air, local markets are the most prominent retail venue in most developing countries, especially for fresh food products (Grace, Roesel, and Lore, 2014). Each local market houses a large number of small-scale retailers operating side by side selling relatively undifferentiated products (Appendix Figure 1). These markets are highly localized and allow for repeated face-to-face interactions between local sellers and consumers.

In such a setting, one would expect the reputation mechanism to be strong and could discipline sellers' behavior. Yet in recent years there have been rising quality complaints of food products sold in these local markets, many in fact stemming from malpractices of the downstream sellers.[14] If we think of quality more broadly as value for money, cheating on quantity is ubiquitous in these markets. Some of these malpractices are easier to detect ex-post than others. The central question is why the reputation mechanism appears to have failed to lead to reliable quality provision, especially given the slow turnover and repeated interactions.

2.2 The Market for Melons

To answer this question, the study focuses on watermelons, one of the most popular products transacted in the local markets and represents 35% of household summer fruits consumption in China (Table 1). Watermelons share many key features with other food products: first, it is a typical experience good for which quality (sweetness) is difficult to detect at the point of transaction. Watermelons are usually sold whole as cut melons are hard to preserve in hot weather. Consumers cannot really detect the fruit's true quality by inspecting the outside (see survey

[13]The online appendix can be found here: https://sites.google.com/site/jiebaiecon/research

[14]For example, formalin-laced tofu, bean cakes, and rice noodles, water-injected pork and poultry, fossil-adulterated flour, etc. An article in *the Guardian* about food safety issues in China: https://www.theguardian.com/sustainable-business/2015/may/14/china-middle-class-organics-food-safety-scares.

evidence below). Second, demand for quality is seemingly high. The best anecdotal evidence is the short-lived popularity of a costly smartphone App developed in 2015 that "claimed" to be able to detect watermelon quality based on the knocking sound. Third, despite the high demand for quality, there is a stark absence of quality premium: sellers in one market all sell one undifferentiated pile of watermelons at the same price.

Quality, on the other hand, varies considerably across watermelons within a given seller's pile. The lightest gray line in Figure 1 plots the cumulative sweetness distribution of 300 randomly picked watermelons from 30 sellers (10 each). 70% of the variation is explained within sellers, suggesting that quality varies tremendously within single batches at each given store. To give a sense of the scale, a sweetness difference of 0.5 matters significantly for taste–sweetness above 10.5 is considered to be very good and that below 9 is very bad. A common saying among Chinese consumers is that "buying watermelons is like buying a lottery–you get a good one if you are lucky". The puzzle is why no seller is providing higher quality (or a better lottery), especially given the opportunity to foster long-term relationships.

Of course, this would not be a puzzle if sellers also could not differentiate quality. Retailers are not growers themselves and most procure their products from the same big wholesale market in the city, where quality is also not differentiated. However, it is well-known that the downstream sellers have some ability to assess quality through inspections of less obvious observables, such as the color of the stripes, sound of knocking, curliness of the veins, etc. These skills are hard to acquire and requires considerable experience to do well. Figure 1 validates this through a sorting test where 30 fruit sellers in the city were asked to sort the 300 watermelons into two quality piles. Details are provided in Appendix D.1. The darkest grey line plots the sweetness distribution of the high pile sorted by the sellers, which statistically dominates the quality of the pool. On the contrary, consumers are unable to assess quality–the lighter grey line plots the distribution for the high pile sorted by 150 consumers, which almost exactly coincides with the unsorted distribution.

It is also worth mentioning that the lack of quality differentiation exists in supermarkets as well. In some other market settings, such as the grocery sector in the United States, stores sometimes carry different reputation and target products of different quality-price levels (e.g. Whole Foods versus Star Market in Boston). Such quality differentiation appears to be absent in the current setting: watermelons sold in the supermarkets in Shijiazhuang are of very similar prices at any given point in time and sometimes even lower quality than those in the local markets because of slower inventory turnover.[15]

Finally, though the study focuses on the downstream markets, one could imagine that if quality can be priced in the downstream, such incentive may trickle up and generate pressure to improve quality for the upstream producers, much as the spillovers via backward linkages (Javorcik, 2004). Currently at the very upstream, farmers have no incentive to control quality because quality is not priced, only weight is. In fact, the wild variation in quality partly originates from bad farming practices that attempt to artificially inflate weight. As we will see, when sellers were induced to provide quality under the experiment, many indeed put more inspection and search efforts into sourcing for better watermelons in the wholesale market.

2.3 Potential Explanations for the Lack of Quality Provision

There may be two potential explanations for the lack of a quality premium at baseline:

A. Low willingness to pay relative to cost

The first explanation is that demand for quality is simply not high enough to cover the effort costs. Consumers' willingness to pay for sweet watermelons is rather small when compared to that for unadulterated milk, uncontaminated meat, or authentic phamaceuticals, the failures of which can lead to health risks. On the supply side, while sellers are very quick at sorting (on average 10 seconds per watermelon during the test), doing that in the hectic wholesale market in early mornings can be mentally stressful, especially under time pressure. Many retail stores are small family businesses and the same person manages both the procuring and selling.

B. Mistrust and imperfect quality control

A different explanation is that mistrust among consumers can make reputation building a difficult and low return investment. A lack of reputation for quality is apparent from baseline consumer surveys. When asked whether any seller in the local market provides higher quality than others, 98% of the consumers answered "No". To quote some respondents, "Everyone behaves badly. It's like buying a lottery no matter whom you go to." Such pessimistic beliefs echo the skepticism about many domestically produced goods, food products and beyond.

[15] In more recent years, we start to see a growing number of fresh fruits retail stores in first-tier cities in China (mostly operating at the neighborhood community scale, replacing the street vendors). Some sell branded fruits at higher prices. However, in most lower-tier cities, the local markets remain as the most common retail venue for daily food products.

The problem is compounded by sellers' imperfect quality control. As we can see from Figure 1, even though sellers sort much better than consumers, their ability to differentiate is also imperfect. This noise in quality control can impede a seller's ability to signal quality. Unless consumers are willing to experiment and upgrade their perception, trust can take a long time to establish.[16]

This paper designs an approach to tease apart the competing hypotheses for low quality provision. The next section develops a model that provides a framework for thinking about quality provision with information problem. The model motivates the experiment and serves as a building block for the empirical model in Section 6.

3 Model: Quality Provision with Information Problem

The basic framework is adapted from Shapiro (1982). The model makes a number of simplifying assumptions about market structure and quality distribution to shed light on three questions: (1) How do sellers pick quality?; (2) How would different prior beliefs affect sellers' decisions about how much effort to put in?; (3) How would demand for quality (relative to cost and quality control technology) affect these decisions?

3.1 Basic Setup

Supply side: A long-run seller faces a fixed pool of consumers. In each period, the seller could choose to sell just one "normal" product, or she could choose to introduce a new "premium" product and sell both. The per-unit cost (P_W) and price (P_N) of the normal product are assumed to be fixed. Let $\underline{\gamma}$ denote the quality of the normal product, where quality is operationalized as the probability that a consumer finds the product satisfactory. In other words, there are only two types of watermelons, good and bad, and each seller carries a mix (recall Figure 1). It is possible to extend the model to continuous quality space and that would not change the main takeaways. Assume that $\underline{\gamma}$ is exogenously fixed and known by consumers.

If the seller chooses to introduce a premium product, she chooses the quality γ_H, which

[16]In other settings, we sometimes observe sellers giving out free samples as a way of signaling quality. For watermelons, since quality varies within single batches, the quality of one is not indicative of the quality of others; it is too costly for sellers to cut open every single watermelon as once open, it is hard to preserve under high temperature.

is initially unobserved by consumers.[17] The additional marginal cost $C(\gamma_H; \underline{\gamma})$ is increasing and convex in γ_H. This cost can be thought of as the inspection effort costs when sourcing watermelons from the wholesale market. The seller also sets the price of the premium product, denoted as P_H^t. For simplicity, to focus on the seller's optimal policies of the premium option, I assume that the price and quality for the normal product are the same as that under no differentiation.

Demand side: There are many ways that one could model consumers' behavior and beliefs when the seller introduces a premium option. The model here focuses on the aspect of consumer learning, which may play an important role for newly introduced experience goods. In this setting, consumers are not informed about the experiment, therefore it is plausible from their perspective to regard the new product as coming from some alternative *upstream source* with some underlying quality that is initially unknown but can be learned over time via actual consumption experiences.[18]

To model the learning process, I adopt a similar framework to that in Dickstein (2014). Suppose that prior beliefs about γ_H follow a beta distribution with parameters (a_0, b_0), where a_0 can be interpreted as the number of prior good experiences and b_0 as the number of prior bad experience. The prior mean is given by $\mu^0 = \frac{a_0}{a_0+b_0}$. Let e_t denote period t's experience realization, which is a Bernoulli random variable with satisfaction probability γ_H. For tractability, I assume that all consumers receive the same experience shock in each period and that information is shared to those who do not purchase by word of mouth.[19] Since beta distribution is the conjugate prior for Bernoulli likelihood, beliefs in period t, after a sequence of experience realizations $\mathbb{e}^{t-1} = (e_1, \ldots, e_{t-1})$, simply follow a beta distribution with parameters $(a_0 + s_{t-1}, b_0 + f_{t-1})$, where s_{t-1} and f_{t-1} are the number of satisfactory and non-satisfactory experiences up to time $t-1$.

In each period, consumers either buy one unit of the product or do not buy any product at all. The utility of not buying is normalized to 0. Consumers' valuation is uniformly distributed

[17] The model analyzes the case of a once-for-all quality choice. In principle, it is possible for sellers to adjust quality and price in every period, however that period is defined. Section 3 of Shapiro (1982) considers such a case and the qualitative conclusions are similar: (1) asymmetric information could lead to quality deterioration and (2) prior beliefs matter for seller's incentive to provide quality.

[18] We can think of the learning dynamics as a reduced form way of capturing learning in a larger Bayesian game in which consumers are trying to infer the *upstream's type*.

[19] In reality, consumers receive different experience shocks. In the structural estimation, I enrich the model by allowing individuals' beliefs to diverge over time with observed experience realizations in the data.

between $[\underline{\theta}, \overline{\theta}]$ with mass M. For a consumer with valuation θ who buys a product at price P, the utility is $\theta - P$ if the product is satisfactory and $-P$ if it is not. In each period, consumers make their purchase decisions to maximize the expected current period utility.

The seller's problem: The seller chooses whether to introduce a premium product, quality and price to maximize discounted sum of profits with discount factor δ. Let $Q^t_{N,\text{nodiff}}$ denote the demand under no differentiation (i.e. just selling a normal product), $Q^t_{H,\text{diff}}$ and $Q^t_{N,\text{diff}}$ denote the demand for the premium and normal products under differentiation.[20] Under no differentiation, the discounted sum of profits are fixed, given by the parameters of the model:

$$\Pi_{\text{nodiff}} = \sum_{t=1}^{\infty} \delta^{t-1}(P_N - P_W)Q^t_{N,\text{nodiff}} \qquad \text{where} \quad Q^t_{N,\text{nodiff}} = (\overline{\theta} - \frac{P_N}{\underline{\gamma}})\frac{M}{\overline{\theta} - \underline{\theta}} \tag{1}$$

Under differentiation, the seller faces a dynamic demand system. In particular, $Q^t_{H,\text{diff}}$ and $Q^t_{N,\text{diff}}$ are functions of $\mu^{t-1}(\mathbb{e}^{t-1}(\gamma_H); a_0, b_0)$, which evolves over time as consumers learn. The expected discounted sum of profits under γ_H is

$$\Pi_{\text{diff}}(\gamma_H) \equiv \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} \max_{P^t_H} \Big(\big(P^t_H - P_W - C(\gamma_H; \underline{\gamma})\big) Q^t_{H,\text{diff}} + (P_N - P_W)Q^t_{N,\text{diff}}\Big)\right] \tag{2}$$

where the expectation is taken over sequences of experience shocks $\{e_t\}_{t=1}^{\infty}$ generated by γ_H. For a given γ_H, the optimal P^t_H is imposed by static profit maximization (because the stylized model assumes complete information diffusion, there is no dynamic implication of current sales). Let γ^*_H denote the maximizer of $\Pi_{\text{diff}}(\gamma_H)$. Finally, suppose there is an initial fixed cost F of introducing a premium option. The seller chooses to differentiate if and only if $\Pi_{\text{diff}}(\gamma^*_H) - F > \Pi_{\text{nodiff}}$.[21]

This completes the setup of the model. In Section 6, I provide supporting evidence on the model's key assumptions and enriches the basic setup by incorporating greater dimensions of consumer heterogeneity, private experience shocks, and market competition.

[20] It is possible that the profit maximization decision is to only sell the premium product. This happens when costs of providing quality is very low. However, such behavior is not observed and I exclude the case here for convenience.

[21] F is not needed for deriving the comparative statics. Without F, if non-differentiation is the optimal strategy under asymmetric information, it is also the optimal strategy under symmetric information and under the first best because only the highest valuation $(\overline{\theta})$ matters for the decision on the extensive margin. However, this is a knife-edge case–any initial costs of introducing the premium product would break it.

3.2 The Effects of Prior Beliefs

The model highlights two broad explanations for the lack of quality provision: (1) high costs of implementing higher quality control relative to demand, and (2) the information problem. First, if cost is high relative to consumers' valuation, for instance $\overline{\theta} < C(\underline{\gamma}; \underline{\gamma}) + P_W$, then higher quality may not be demanded and supplied even under symmetric information. Second, since a seller's claim of offering high quality cannot be immediately verified, consumers' prior beliefs matter. Sellers who rationally discount future profits may lack the incentive to provide quality if beliefs are pessimistic and trust takes a long time to establish. Hence, markets can get stuck in a self-confirming equilibrium with no quality provision.

In reality, these two aspects act jointly. However, the welfare implications are very different: under the former, the distortion caused by the information problem is small, whereas under the latter it could be large. In practice, it is hard to directly infer and estimate sellers' inspection cost function. To understand the main barrier for quality provision, the experiment aims to create exogenously variations in prior beliefs. These variations should have minimal effects if the key barrier for quality provision is high costs. On the other hand, if the information problem is the key barrier, enhancing prior beliefs could significantly strengthen sellers' incentives to provide quality. The effects are stated in the following two propositions:[22]

Proposition 1: *(Incentive to differentiate)* $\Pi_{\text{diff}}(\gamma_H)$ increases with a_0 and decreases with b_0.

Enhancing prior beliefs, either by increasing a_0 or decreasing b_0, raises the seller's return under differentiation. The intuition is straightforward. Holding a_0 fixed, a lower b_0 implies a higher prior mean and a larger prior variance, and hence a faster speed to establish trust and larger discounted returns. The next proposition examines how the seller's optimal quality choice responds to prior beliefs if she differentiates.

Proposition 2: *(Optimal quality choice)* If $\frac{\partial^2 \Pi_{\text{diff}}(\gamma_H^*; a_0, b_0)}{\partial \gamma_H \partial a_0} > 0$, γ_H^* increases with a_0. Similarly, if $\frac{\partial^2 \Pi_{\text{diff}}(\gamma_H^*; a_0, b_0)}{\partial \gamma_H \partial b_0} < 0$, γ_H^* decreases with b_0.[23]

Proposition 2 states a simple comparative statics. If prior beliefs and quality are comple-

[22]There is a direct mapping from (a_0, b_0) to the mean and variance of a beta distribution: $m = \frac{a_0}{a_0+b_0}$; $\text{var} = \frac{a_0 b_0}{(a_0+b_0)^2(a_0+b_0+1)}$. One can write down similar predictions in terms of the latter. I stick with (a_0, b_0) to be consistent with the structural analysis, which directly estimates a_0 and b_0 for different branding technologies.

[23]In general, γ_H^* is non-monotonic in a_0 and b_0. When $a_0 + b_0$ is sufficiently large, as one of the two parameters tends to 0 (very pessimistic or very optimistic beliefs), the incentive to provide quality vanishes.

mentary, the seller will provide higher quality as prior beliefs improve.[24]

The experiment exogenously varies consumers' prior beliefs via different branding technologies to test the two predictions. The next section describes the design and data.

4 Experimental Design and Data Collection

4.1 Experimental Design and Timeline

The experiment was conducted in Shijiazhuang, China.[25] The city has over 800 gated communities and more than 200 local markets. Randomization happened at the market level. 60 sellers located in 60 different markets were recruited to participate in the study following an initial screening procedure to minimize heterogeneity in the study sample for power concerns and logistical purposes. Details for the screening process and selection criteria are described in Appendix C.1. Sellers were paid a fixed amount of 100 RMB/week for taking part in the study, which mainly compensates for their time to record daily sales data (Section 4.2).

All 60 sellers signed an agreement at baseline that they would experiment with quality differentiation for the first two weeks, that is, selling two piles of watermelons: a premium pile and a normal pile. It was made clear to the sellers that the research team would not interfere in any other aspect of their business, in particular they were free to choose the quality, price, and quantity for each pile.

The 60 sellers were randomized into 6 groups:

Branding treatments. Sellers were randomized into one of the three branding groups: laser, sticker and label-less. Every morning, surveyors visited the sellers' stores and performed a free branding service. For the laser group, surveyors used a laser-engraving machine to print a laser-cut label of the words "premium watermelon" ("Jing Pin Xi Gua" in Chinese Pinyin) on the watermelons in the premium pile. For the sticker group, surveyors pasted a sticker with the same words. For the label-less group, surveyors did nothing. Note that the branding treatment was only for watermelons in the premium pile, picked by the sellers themselves; those in the normal pile were left as they were. Figure 2 shows pictures of the branding treatments. Most

[24]In reality, optimistic beliefs also encourage sales, enabling information to spread faster, rewarding good behavior and punishing bad behavior faster. The channel is absent in the stylized model (with perfect information diffusion) but will be captured in the empirical model.

[25]Urban area: 154.2 sq mi; urban population: 2,861,784; urban density: 19,000/sq mi

sellers sold two piles of watermelons at the beginning of the experiment, but some reverted back after some time. For those sellers, the branding service was withdrawn because there was no longer a premium pile.

Compared to stickers, laser is a new and "uncontaminated" branding technology. Because the cost of laser machines is very high ($\approx$ 8k USD), consumers generally regard laser branding as more effective at deterring counterfeits than stickers, which can be cheaply fabricated and are highly "contaminated" by rampant counterfeiting activities in the past.[26] This is also true in other developing country contexts where brand protection is weak (Qian, 2008). In this setting, laser could potentially wipe out bad historical stereotypes associated with stickers (lower b_0), thereby allowing trust to establish faster. Proposition 1 and 2 suggest that sellers in the laser group may have a stronger incentive to differentiate and provide higher quality.

There could be other ways in which laser branding affects market outcomes. For example, it may simply represent something "cool" and directly affect utility other than through signaling quality. Alternatively, it may lead to more optimistic beliefs because it is expensive (the argument, known as forward induction (Klein and Leffler, 1981), is formalized in Appendix B). However, these alternative stories do not explicitly incorporate the role of learning and trust-building dynamics. I address these alternative interpretations in Section 5.5 and provide evidence that supports the learning model.

A cross-randomized incentive treatment. Within each branding group, half of the sellers were randomly given an incentive to maintain quality for the premium pile. The incentive treatment was enforced via unannounced quality checks twice per week. At every check, surveyors randomly picked one watermelon from the premium pile and one from the normal pile. The quality of both was measured using a sweetness meter (Appendix Figure 2). For sellers in the incentive group, if the sweetness of the former attained 10.5 at both checks, sellers received a monetary reward of 100 RMB at the end of the week, on par with daily sales profits. Sellers in the non-incentive group received the same quality checks, but were not given any reward. The incentive was removed in the later part of the intervention, and that was unanticipated by the sellers.

The incentive treatment represents a way of subsidizing seller's initial trust building. We

[26] In a consumer pilot survey, consumers were asked about their willingness to pay for watermelons under different branding technologies. The reported willingness to pay for laser-branded watermelons is significantly higher than that for sticker-branded ones, and the primary reason cited is that the former is more likely to be "authentic" whereas the latter can be highly adulterated by counterfeits.

can think of it as shifting the posterior beliefs: if the incentive could motivate sellers to provide higher quality, then upon its removal after some period T, sellers who have had the incentive are essentially endowed with higher reputation than those sellers who have not had the incentive. Proposition 2 suggests that higher quality may be sustained even in the post-incentive period. This, therefore, provides a further test for the model, and also helps to identify sellers' effort cost in the structural estimation.

Summary. In total, there were 6 distinct treatment units. Randomization was stratified on housing prices, a proxy for local income and potential demand for quality. Appendix Figure 3 shows a map of the 60 sellers, marked by groups. Note that these markets are geographically segregated (average distance between two closest markets is about 1 kilometer). Since watermelon transactions are highly localized, spillover effects across markets should be minimal.[27] However, there could be spillovers to the non-sample sellers operating in the same 60 markets. On average, each market houses 3 fruit sellers (Table 1) and only 1 was in the study sample. I collected data on the other sellers' pricing and differentiation behavior to examine any strategic responses. The demand side data (see below) further allows me to quantify the magnitude of market share reallocation.

The main analysis focuses on the 60 sellers in the study sample. Since the label-less non-incentive group is subject to the same nudge for quality differentiation, monetary reward for recording data and random quality checks as the other five groups, differences in outcomes across groups isolate the branding and the incentive effects as well as their interactions.

Timeline. Figure 3 describes the timeline. The market intervention was rolled in from July 13 to July 19, 2014. Two weeks into the intervention, an announcement was made to all sellers that they were free to decide whether they wanted to continue with quality differentiation or not. This allows me to examine differential incentives across groups. Six weeks into the intervention, the incentive was removed. The intervention was phased out from September 6 to September 12. An endline survey was conducted upon the surveyors' final visit to sellers' stores, and two follow-up surveys were conducted to examine longer-term outcomes.

[27] From the baseline survey, 80% of watermelons are bought from a given household's local market; the remaining 20% are bought from nearby supermarket rather than other local markets. During the experiment, I did not observe many consumers switching between markets (for example, from a label-less market to a laser market).

4.2 Data: Overview

Baseline surveys. Table 1 summarizes the sample characteristics. Most sellers sell fruits all year long and do not expect to relocate. The median household consumes 1 watermelon per week in the summer, and 75.6% list the local market as the main source of purchases.

Supply side: quality, prices, and sales. Quality data (measured in sweetness) were collected from the biweekly random quality checks as described in Section 4.1. One concern is that sellers and surveyors may collude to manipulate the outcomes, especially for those in the incentive group. To minimize this concern, surveyors were rotated across markets on a weekly basis. Surveyors' also collected daily retail prices for both the sample sellers and the other sellers in these markets, as well as the daily wholesale price. Sellers were asked to record their daily sales for watermelons and peaches by quality category (Panel A of Appendix Figure 4). In total, there were 60,806 transaction records over the course of the intervention. 81% of transactions were for watermelons and 19% were for peaches. On average, sellers sell 257 Jin ($\approx$ 340 pounds) of watermelons per day, and the average daily sales profit is 103 RMB. Here and in all subsequent regression analysis, sales profit is computed using sales quantity and the difference between retail and wholesale prices. Therefore, it does not take into account any transportation/storage costs and effort costs of sourcing for higher quality.[28] For most of the analysis, I aggregate the transaction-level sales to seller-day-quality category level.

Demand side: household panel purchasing. 675 households in 27 communities, evenly distributed across the treatment groups, were recruited to record the family's entire summer fruit consumptions (Panel B of Appendix Figure 4). For each purchase, households were asked to record the date of the purchase, the place of the purchase, the quantity bought, the amount paid, whether the purchase was made from the sample seller or from other places (including other sellers in the local market), and whether the purchased fruit had any branding on it or not. Importantly, households were asked to report a satisfaction rating ranging from 1 to 5, where a higher number indicates higher level of satisfaction. This provides a second measure of quality, besides sweetness, which does not subject to the concern of collusion. In total, there were 15,292 purchase records, of which 30.8% were for watermelons. The median for the number of watermelons consumed per week is 1 and the mean is 1.15 with standard deviation 1.06. These

[28] Specifically, sales profit = premium pile price $\times$ premium pile sales quantity + normal pile price $\times$ normal pile sales quantity - total sales quantity $\times$ wholesale price. Alternatively, I can use the recorded sales values to calculate profits, which takes into account consumer bargaining. Results are qualitatively robust.

numbers match well with Table 1.

Endline and follow-up surveys. The seller endline survey was conducted during the surveyors' final visit to the sellers' stores and elicited sellers' willingness to pay for different branding technologies. The household endline survey was distributed together with the last week's recording sheet and elicited households' willingness to pay for quality under different branding technologies. Two follow-up surveys were conducted one week and one year after the intervention to examine longer-term behavior. Attrition rate is small: 1 seller dropped out during the intervention because the market was closed for road construction. For the second follow-up, surveyors were able to locate 57 of the original 60 sellers.

Details of the recruiting procedure, data collection, and issues with cleaning the sellers' and households' recording data were discussed in detail in Appendix C.1-C.4. Balance checks on market, seller and household baseline characteristics are in Appendix Tables 1 to 3.

5 Experimental Evidence

This section presents experimental evidence on the effects of the branding and the incentive treatments. Figure 4 plots the number of sellers who differentiated quality at sale in each treatment group over time. We see that sellers in the label-less groups sharply reverted back to baseline after the first two weeks. On the other hand, most sellers in the sticker and laser groups continued to differentiate till the end of the intervention.

The rest of the section examines demand side's responses, sellers' quality, pricing and sales in order to understand the differential incentives, in particular, why sellers who were induced to differentiate during the intervention had not already done so at baseline.

5.1 How Do Different Branding Technologies Affect Prior Beliefs?

First, are consumers' initial beliefs less stigmatic under laser than under sticker? This question is difficult to examine in a regression framework as beliefs are not directly observable. To answer this question, I exploit the household panel data where I observe both the purchase decisions and the self-reported satisfaction rating for each consumption experience.

Table 2 provides some suggestive evidence. The data is collapsed to the household-week level. The dependent variable is a dummy for whether a household purchased any premium watermelons in a given week, regressed on two measures of past experiences: (1) the average lagged satisfaction rating of all past premium purchases, and (2) the percentage of past purchases that attain the highest rating. Note that if a household has never purchased any premium watermelons in the past, these measures are not defined. Therefore, the coefficients are estimated from household-week observations conditioning on a positive number of premium purchases prior to a given week.

Column 1 and 2 of Panel A show that lagged experiences strongly predict repurchasing decisions for households in the laser markets. To interpret the magnitudes, take the estimate in column 2, which shows that for two similar households at a given point in time, the household that has had only very good past experiences is 45% more likely to repurchase a premium watermelon than the household that has not had any very good experiences (but that has experienced the product). On the other hand, the coefficients are much smaller and less precisely estimated for households in the sticker groups, as shown in columns 3 and 4. These patterns are consistent with a learning framework and suggest that prior beliefs may be more "stubborn" under stickers, which implies that purchasing decisions would be less responsive to past consumption experiences.

Panel B repeats the same exercise for purchase decisions of the normal pile. Since consumers have experienced unlabeled watermelons for a long time, each additional experience should weigh less. Indeed, the coefficients are small and insignificant.

5.2 How Do Different Branding Technologies Affect Quality Choice?

Do the different learning dynamics affect sellers' incentives to investing in quality? Panel A of Table 3 compares the premium pile quality, measured in sweetness, for sellers in the sticker and laser groups. Standard errors are clustered at the seller (market) level, which is the unit of randomization. This applies to all the regression analysis below unless otherwise stated. To address the concern of a relatively small sample size, I also conduct two small-sample robustness checks using permutation test (Bloom, Eifert, Mahajan, McKenzie, and Roberts, 2013) and clustered bootstrap (Cameron, Gelbach, and Miller, 2008). The p-values are reported in the table. The results are consistent with Proposition 1 and 2: on average, sellers in the laser group provide significantly higher quality than sellers in the sticker group, with and without the incentive. The same pattern holds when using household satisfaction rates as a measure for quality (Appendix Table 4).

To further understand sellers' behavior, I look at how the quality of the premium pile compares with that of the normal pile and how the two compare with the market average. To focus on the effect of the branding treatment, I restrict the sample to sellers in the non-incentive groups. Panel A of Table 4 shows that the average quality of the premium pile is significantly higher than that of the normal pile. However, the difference could be either due to a genuine quality improvement of the premium pile or a deterioration of the normal pile. To examine the latter possibility, Panel B runs the same regression but with quality difference from the market average as the outcome variable. I use the average sweetness of randomly picked watermelons from sellers in the label-less group after they had reverted back to non-differentiation as a proxy for the average quality of the pool.[29] Column 3 shows that sellers in the laser group maintained a higher quality for the premium pile and kept the normal pile quality on par with the market average. This suggests that sellers may have put more effort into sourcing good watermelons. Qualitative evidence from a followup survey supports this finding: when asked to describe their activities in the wholesale market during the intervention period, sellers in the laser group reported more time spent in the wholesale market sourcing for watermelons, and they were also more likely to shop around wholesalers (Appendix Figure 5).

The evidence for the sticker group is quite mixed. On average, the quality of the normal pile appears to be lower than the market average and the quality supremacy of the premium pile is not significantly different from 0 (p-value of 0.584). The large standard errors indicate considerable heterogeneity across sellers in the sticker group. Anecdotally, some sellers in the sticker group simply labeled all watermelons except for a few observationally bad ones, which they then marked down and sold as a low-end product. While the sample size is too small to formally examine heterogeneity within a treatment group, I note the difference between the sticker group and the genuine quality-price premium observed for the laser group.

5.3 How Do Different Branding Technologies Affect Sales?

Table 5 examines the effects of the branding treatments on sales outcomes. The outcome variables are at the seller-day level, including log sales profits (in RMB), price premium above market average price (in RMB/Jin), sales quantity (in Jin) for each pile, and the total sales

[29] Watermelon quality could fluctuate over the season. Therefore, this time-varying measure of average quality is better than a fixed baseline measure.

quantity.[30] If a seller stops differentiating quality, the unit price for the premium pile is defined to be the same as that for the normal pile and sales quantity for the premium pile is coded as 0. All regressions include day fixed effects (λ_t) to control for time-specific aggregate shocks, such as weather. The even columns control in addition for community and seller baseline characteristics.

Column 1 and 2 show that on average, the laser group earns 30-40% higher sales profits than the label-less group. This is due to both a higher price (columns 3 and 4) and higher sales quantity for the premium pile (columns 5 and 6). Sales of the normal pile are not significantly different from the label-less group. The results suggest that sellers in the laser group attract more high-end customers without losing sales on the normal product. Other competitors in the market did not mimic the sample sellers and differentiate quality at sale as they were not given access to the technology. I also do not observe significant strategic pricing responses among the other sellers in the laser markets (Appendix Table 5).

On the other hand, for the sticker group, sales of the premium pile appear to be lower on average than the laser group (the p-value of a one-sided test is 0.238) despite a lower price. Furthermore, sales of the premium pile (columns 5 and 6) are offset by a reduction in the sales of the normal pile (columns 9 and 10). As a result, total sales and profits are not significantly different from those of the label-less group which reverted back to baseline.

These results explain why sellers did not differentiate quality at baseline despite the fact that stickers have long been cheaply available. For the laser group, the relevant consideration is whether the increase in sales profits, netting out effort costs of providing higher quality, justifies the fixed cost of the laser machine. I come back to this point in Section 5.6.

5.4 How Does the Incentive Treatment Affect Quality Choice?

Panel B of Table 3 shows that the incentive did lead sellers to provide higher quality for both sicker and laser groups. To examine whether higher quality was sustained in the post-incentive period, Table 6 runs a difference-in-difference regression. The coefficient for the interaction term between the incentive treatment and the post-incentive dummy is close to zero and not significant for the laser group. On the contrary, sellers in the sticker incentive group seemed to revert to a lower quality level after the incentive was removed. These results are consistent with

[30] Here and in all subsequent analysis with prices, I use the listed prices as observed by surveyors during the morning visits to the markets. Alternatively, I can use the effective prices, calculated as total daily sales revenue divided by total daily sales quantity for each quality category. Results look very similar.

the conjecture that it may take longer to establish trust under stickers, and thus it is not clear how much the incentive has facilitated initial learning during this short intervention.

5.5 Alternative Models

The experimental findings are consistent with the predictions of the model in Section 3. Appendix D.2 and D.3 present further corroborating evidence exploring the time dynamics of sales and household perception elicited at endline. The differential time dynamics of sales within a branding group (laser incentive v.s. laser non-incentive) cannot be explained by a fixed "laser-coolness" effect, but rather indicate that higher quality only pays off over time (Appendix Table 6). The different endline perceptions among households across markets, after real consumption experiences, also support the learning mechanism, as opposed to a coordination mechanism under perfect public monitoring (Appendix Table 7).

It is also important to acknowledge the role of relationships, which commonly exist in these markets (Fafchamps, 2002). For example, sellers can pick higher quality watermelons for sale to repeat customers. The lack of explicit quality differentiation at baseline would not be a problem if relational contracting has *perfectly* allocated high quality watermelons to high valuation customers. However, if that were the case, we would not expect to see the positive effect on sales for the laser group. This would be true unless there are important spillovers across a seller's multiple products. For example, sellers may use watermelons to build relationships and maintain businesses for other fruits they sell. To examine this possibility, I collected data on sellers' peach sales, another popular summer fruit. I do not find significant differences across the treatment groups in terms of peach sales, at least during this short intervention period (Appendix Table 8). To the extent that sellers' preferential treatment may not perfectly align with people's willingness to pay and there are in fact substantial switchings across sellers among households,[31] there could still be important welfare loss due to allocative inefficiency.

5.6 Interpretation of the Experimental Findings

The experimental findings demonstrate that reputational incentives are present and can be made to pay. As long as providing higher quality involves positive efforts, in a one-shot game, sellers would not exert such additional efforts and would randomly label some watermelons as

[31] Only 1 out of 675 households in the sample reported all of its fruit purchases from a single local seller.

"premium" and sell them at a higher price. Having said that, one year after the intervention, when surveyors revisited these markets, none of the 57 sellers that could be tracked continued with quality differentiation. This suggests individual sellers would not have the incentive to take up this new technology themselves. When asked about their willingness to pay for the laser branding service (if someone could provide it again in the next season but no longer for free), the average reported WTP among the laser group is 0.15 RMB per watermelon (relative to an average gross markup of 4 RMB–see Appendix Table 9). However, this amount multiplied by each seller's small sales volume falls far short of the cost of the laser machine ($\approx$ 50k RMB). Overall, these experimental findings provide a qualitative explanation for the lack of quality provision at baseline.

Two important questions remain: first, since sellers would not fully internalize the gain in consumer surplus, would it be beneficial for a third-party to subsidize the initial trust building process? To answer this, we would need to know the underlying demand and cost parameters and how big the distortion caused by the information problem is in the first place. Second, laser branding is one extreme technology that worked–could something else work as well? The key to answer this question is to understand exactly why laser worked: did it work mainly because (a) the signal is expensive and hence induces a more optimistic prior, (b) it is new and hence learning is more salient, or (c) it is accessible only to a small number of sellers, who could use it to distinguish their products (even with low and stubborn initial priors)? Answering these questions requires going beyond the experimental data. In the next section, I use a structural approach to recover the underlying beliefs, demand and cost parameters. After that, Section 7 uses the structural estimates to study the welfare consequences under various informational environments, market structures and policy counterfactuals.

6 An Empirical Model of Learning & Quality Provision

The empirical model follows the same setup as the model outlined in Section 3. The experiment introduces random variation into a household's choice set: a new "premium" option under different branding technologies. The randomization of the branding technologies combined with observed panel purchase decisions conditioning on realized consumption experiences allows me to recover prior beliefs and willingness to pay for quality. The supply-side parameters are estimated by solving for the sellers' optimal policies, taking the demand estimates as given.

I first augment the setup in Section 6.1 and then describe the two-step estimation procedure in Section 6.2. Section 6.3 discusses the results and examines model fit. Section 6.4 uses the structural estimates to simulate households' beliefs and sellers' net returns evolution under different treatments.

6.1 Setup and Assumptions

6.1.1 Demand Side

The prior distribution and the updating process are described in Section 3. To better fit to the real setting, I enlarged households' choice set to include buying from other sellers and allow beliefs to diverge over time by incorporating private experience shocks. Let $e_{imjt} \in \{0, 1\}$ indicate whether a type j watermelon is satisfactory or not for household i in market m at time t, where $j = 1$ indicates the premium pile from the sample seller, $j = 2$ indicates the normal pile from the sample seller, and $j = 3$ indicates those from all other sources (including other sellers in the same market). Applying the Bayesian updating formula, consumer i's posterior for the quality of the premium option at time t is given by $(a_{im1t}, b_{im1t}) = (a_0 + s_{im1t}, b_0 + f_{im1t})$, where s_{im1t} and f_{im1t} are the numbers of satisfactory and non-satisfactory experiences household i has had till time t. For simplicity, I assume that households do not update on the other options. This is consistent with the reduced form results in Panel B of Table 2 (discussed in Section 5.1). Let q and $q + \Delta q$ denote the (degenerate) beliefs about the quality of other sources and the normal option. In particular, Δq captures the possibility that consumers may downgrade the quality of the normal pile if they suspect sellers pull out the high quality ones. I estimate separate belief shifters, $\Delta q(s)$ and $\Delta q(l)$, for the normal option for sticker and laser groups.

I enrich the basic model to incorporate richer dimensions of household heterogeneity. Expected utility of purchasing option $j \in \{1, 2, 3\}$ at time t is

$$u_{imjt} = (\theta_0+\theta_1\text{WTP}_i)\mu_{imj,t-1} - (\alpha_0+\alpha_1\text{Highinc}_i)P_{mjt} + \beta\text{Num}_i + \eta_i\mathbb{1}_{(j=1)} + \xi_i\mathbb{1}_{(j\in\{1,2\})} + \lambda_m + \lambda_t + \epsilon_{imjt}$$

where $\mu_{imj,t-1}$ denotes household i's posterior for option j at the end of time $t-1$. P_{mjt} is j's price in market m at time t. θ captures vertical taste differentiation, and is allowed to vary across households with different baseline self-reported willingness to pay for quality. The price coefficient α is allowed to be different for high- and low-income groups. Num_i is the number of watermelons consumed per week reported at baseline, which seeks to capture heterogeneous love for watermelons in general. η_i and ξ_i are unobserved preferences for the premium option and for the sample seller (horizontal taste differentiation). λ_m are market fixed effects, capturing time-invariant differences across markets. λ_t are time fixed effects, capturing aggregate time shocks that affect all markets, such as weather shocks. ϵ_{imjt} are idiosyncratic random utility shocks, which are realized in each period before the purchasing decision is made. Let V_{imjt} denote the mean utility, excluding the random shock component.

There is an outside option with mean utility 0 for not purchasing any watermelon in a given period (denoted as $j = 0$). Household chooses j with the highest expected utility. Assuming that the idiosyncratic shocks ϵ_{imjt} follow i.i.d. Type 1 extreme value distribution, the choice probability takes a logit form:

$$\text{Prob}_{imjt} = \frac{\exp(V_{imjt})}{\sum_{k=0}^{3} \exp(V_{imkt})}$$

The current model abstracts away from forward-looking behavior, which can be important for experience goods. The short panel of the purchasing data is limited in addressing this aspect and the empirical evidence is mixed.[32] The goal of the structural exercise is to estimate a parsimonious model that can describe the observed purchasing behavior, and that is also tractable enough to be integrated with the supply side.[33]

6.1.2 Supply Side

For the supply side, I focus on the laser groups, for which we have seen clear evidence for providing quality. Sellers choose prices and quality to maximize the net present value of profits (Equation 2). I parameterize the additional marginal cost of providing the premium product $C(\gamma_H)$ as $c\log(\frac{1-\underline{\gamma}}{1-\gamma_H})$, where $\underline{\gamma}$ denotes the average quality of the undifferentiated pool. $C(\gamma_H)$ captures the effort costs of sourcing better watermelons in the upstream. In the extreme case, if $\gamma_H = \underline{\gamma}$, the cost simply reduces to 0. Finally, to match the empirical setting, the objective function for the incentive group contains an additional term $\phi(\gamma_H^t)B$ capturing the expected

[32] Given the seasonal nature of the fruit, if the option value of experimentation plays an important role, we may expect that the number of first-time buyers for the less-known premium product option to be higher in the initial period (within the current model). There does not appear to be such a pattern in the data.

[33] To model forward-looking behavior, besides the usual computational difficulties of solving a dynamic discrete choice problem (discussed in detail in Ching, Erdem, and Keane (2013)), the current setting poses an additional challenge, which is that it may be hard to model the value of experimentation in the context of a new good as households' perceptions about future product availability, price and quality would matter.

incentive payment, where $B = 100$ RMB and $\phi(\gamma_H) = \gamma_H^2$ (since quality checks were conducted twice per week).[34]

The main estimation challenge for solving the dynamic optimization problem is that the seller's state space, which is the joint distribution of household beliefs (μ^t) and characteristics (X) included in the demand model, is of infinite dimension. To make progress, I make an important simplifying assumption that seller pegs the normal pile price at the market average in each period and chooses a once-for-all quality (γ_H) and price premium (m_H) for the premium pile (i.e., $p_H^t = p_N^t + m_H$). Appendix Figure 6 and 8 plot the price and quality trajectories for the laser groups. We do not observe a clear time pattern. Appendix Table 10 further examines the time dynamics in a regression framework, and the coefficients for the time variables are very close to zero.

The empirical patterns provide some qualitative justification for the assumption. One possible explanation is that frequent price adjustments may send some negative signals to consumers, and although quality differentiation happens daily, to actually fine-tune that to actual demand conditions may be hard and mentally costly. Having said that, a seller may well increase price in longer-time horizons as beliefs evolve. Unfortunately, the data, which only lasts for 8 weeks, is limited in addressing these long-term dynamics. Given this limitation, the approach taken here searches for the optimal policy within the restricted class of policies.

6.2 Estimation and Identification

6.2.1 Demand Side: Simulated Maximum Likelihood Estimation

The demand model is estimated using simulated maximum likelihood (Train (2009)). I collapse the household panel purchasing data to household-week level and merge that with the market-week level average prices. For each purchase experience, the household reports a satisfaction rating from 1 to 5. I recode 5 to be satisfactory and $\{1, 2, 3, 4\}$ as well as missing values to be non-satisfactory. I estimate separate prior beliefs for laser and sticker. We can think of households living in different markets as facing different choice sets: households in the laser and sticker markets face a premium option with either a laser or a sticker label. For households in the label-

[34] The implicit assumption is that the pre-specified sweetness threshold matches consumer's subjective satisfaction assessment. Appendix Figure 9 plots the empirical CDF of the sweetness for the undifferentiated piles. 10.5 corresponds to the 73rd percentile of the distribution, which is close to the 30% empirical satisfaction rate in the household data for undifferentiated watermelons.

less markets, they face a restricted choice set without the premium option (from week 3 onwards). Occasionally there are multiple purchases for a household in a given week. This is accommodated by applying the Bayesian updating formula multiple times based on all the realized experiences in that week. Details for the estimation procedure and standard error calculation are provided in Appendix E.1. Appendix E.2 discusses alternative prior specifications, including a Dirichlet's prior and different rating thresholds.

The identifying assumptions are twofold: first, market and time fixed effects fully capture unobserved time-varying shocks that directly affect both price and demand for a market. Second, η and ξ fully capture unobserved persistent household heterogeneity. Under these assumptions, with one period data on market shares, we can identify the market specific constants, the mean of the prior beliefs multiplied by the vertical taste parameters, the price coefficients, the coefficient for *Num*, and the distributions of η and ξ (following standard arguments in the discrete choice literature). Parameters θ, a_0 and b_0 are identified from the dynamic purchasing patterns. Intuitively speaking, if repurchasing decisions are very responsive to past experiences, it could either because households care a lot about quality (large θ) or the variance of the prior is large (small a_0 and b_0). However, the *difference* in the *change* in the repurchasing probabilities between going from zero to one good (or bad) experience and that going from one to two separately identifies these parameters. In particular, the difference should be bigger under the large variance story than it is under the large willingness to pay story because belief updating is more salient in the former case. Appendix Table 11 summarizes the repurchasing probabilities conditioning on experiences. The patterns are largely consistent with the results in Table 2.

6.2.2 Supply Side: Minimum Distance Estimator

Taking the demand estimates as given, the supply side parameters are estimated using a minimum distance estimator. Ideally, I would like to solve for the optimal policies market by market and apply the minimum distance estimator to the full vector of policies for all sellers. Unfortunately, γ_H is not observed for each individual seller and cannot be reliably approximated using the empirical satisfaction rate due to the small household sample size for each market. Given this data limitation, I first construct a hypothetical *average market* by pooling together all households in the laser markets and averaging the market fixed effect estimates. I then solve for the optimal policies, m_H^* and γ_H^*, for a seller facing this hypothetical market. The structural parameters are estimated by minimizing the distance between the optimal policies and the empirical average policies. Details of constructing the hypothetical market and the empirical average policies are discussed in Appendix E.3.

For each given set of δ and c, the optimal policies are found using grid search. The objective function is minimized by searching over grids of δ and c. Intuitively speaking, low quality provision could be either due to high costs or low discount factors, but the former implies a larger quality gap between the incentive and non-incentive groups: the more convex the cost function (larger c), the steeper the increase in the costs of improving quality, which dampens the effect of the incentive.

6.3 Results and Model Fit

The simulated maximum likelihood (ML) estimates are presented in Table 7. Market and time fixed effects are reported in Appendix Table 12. To match the small market shares of the premium option in the first week, I constrain the prior mean (a_0) to be zero in actual estimation (see discussion in Appendix E.1).[35] Estimates of other key parameters are qualitatively similar to the unconstrained case.

Looking at column 1, the estimate for b_0 is 0.938 for laser and 2.578 for sticker. The point estimates are consistent with the reduced form results in Table 2 and suggest that prior beliefs are more *stubborn* under sticker than under laser. In particular, one satisfactory experience updates the posterior mean to 0.52 under laser, but only to 0.28 under sticker. Thus the different estimates of b_0 imply very different belief updating behavior and capture different dynamic purchasing patterns (despite the constraint on prior mean). Beliefs about the quality of the undifferentiated option from the other sellers is estimated to be 0.307. This number matches well with the 30% empirical satisfaction rate in the household data. The negative $\Delta q(s)$ suggests that consumers in the sticker markets seem to perceive the normal pile as having lower quality if sellers sell it beside another pile that is labeled with a sticker. This is in fact consistent with sellers' actual behavior shown in column 4 of Table 4. Signs of the other estimates are consistent with expectations.

Appendix Figure 10 and 11 examine model fit by looking at the dynamics of market share

[35] For Beta distribution, a_0 and b_0 need to be positive. However, the unconstrained ML estimation hits the boundary for a_0, and this is because of a combination of: (1) very low market share of the premium ptopion in the first week (suggesting a low prior mean) and (2) very responsive subsequent purchasing conditioning on realized experience (suggesting a high prior variance). I discuss this issue and alternative parametrization and estimation strategy more in Appendix E.1.

and repurchasing probabilities conditioning on experiences. Overall, the purchasing patterns generated by the prior estimates and the Bayesian learning process mimic the actual purchasing patterns well.

Columns 2 to 4 of Table 7 consider three extensions to the baseline model by considering direct utility of laser, correlated learning, and information diffusion. Details are described in the table note. Overall the ML estimates stay quite robust across various specifications and the likelihood ratio test does not reject the baseline model. A static model without learning, on the other hand, is rejected by standard information criterion test (not reported).

Taking the demand estimates in column 1 of Table 7, δ and c are estimated to be 0.98 and 0.64.[36] The model is able to generate a quality gap between the incentive and non-incentive groups (0.48 versus 0.41), which is fairly close to the empirical gap (0.53 versus 0.40). The optimal price premium for the incentive group is also higher than that for the non-incentive group, though the magnitudes are larger than the empirical values. Table 8 simulates aggregate sales outcomes using the parameter estimates and the average empirical policies. Overall, the simulated weekly average sales quantity and profits are in line with the actual sales outcomes in the data.

6.4 Beliefs and Net Returns Evolution

The estimates in Table 7 suggest that laser branding worked mainly because of a higher prior variance, which makes learning more salient, rather than a higher prior mean. I now use the structural estimates to examine how beliefs endogenously evolve over time, and how prior beliefs affect seller's incentive to provide quality.

Panel A of Figure 5 plots the market average beliefs evolution for the quality of the premium pile. We see that the average beliefs are the highest for the laser incentive group by the end of the intervention. Conditioning on the incentive treatment, average beliefs rise faster under laser than under sticker. This is a result of two underlying effects: first, laser branding induces faster belief updating; second, laser branding induces sellers to provide higher quality, resulting in more satisfactory experiences. To decompose the two effects, Panel B simulates counterfactual beliefs evolution under three scenarios: (1) sticker prior and sticker group's empirical policy (dashed line); (2) laser prior and sticker group's empirical policy (dotted line with square markers); (3)

[36] Appendix Figure 12 plots the value of the objective function as δ and c vary and Appendix Table 13 reports the optimal policies under various δ and c in comparison with the empirical policies.

laser prior and laser group's empirical policy (dotted line with diamond markers. Comparing (1) and (2), we see that holding the supply-side behavior as fixed, laser branding alone has a significant impact on beliefs evolution. This difference shape sellers' incentives to provide quality, which further drive markets to different outcomes over time. The gap (1) and (3) represents the total effect.

Figure 6 plots seller's net profits evolution. An extrapolation to 5 seasons suggests that there might be large gains under laser (Table 8): the five-season discounted sum of net profits is ≈ 13 kRMB higher than baseline (≈ 11 kRMB).[37] However, this increase is still not large enough to justify the initial investment cost of the laser machine (≈ 50-60k RMB) for small individual sellers. This rationalizes the experimental findings. One idea is that sellers in one market could collectively adopt the new technology to spread the fixed cost. However, markup may also be lower in that case due to competition, which could diminish the incentive to provide quality. In other words, it could be that laser works precisely because it is only accessible to a few. This discussion points to the importance of understanding the role of fragmented market–a common feature of many industries in developing countries–in the presence of information problems. I turn to this in the last section.

7 Welfare of Information Friction & Fragmented Market

In a second-best world with multiple frictions, the welfare implication of each friction is theoretically ambiguous as the different friction could counteract. In particular, while market power generally distorts quality provision from the first best (i.e., Spence distortion), it also internalizes the return of investing in quality by allowing sellers to capture a larger portion of the consumer surplus. To examine the interaction, I conduct counterfactual exercises that remove one imperfection at a time in order to isolate the effect of the other. These exercises involve extrapolation beyond the sample period and additional assumptions on conducts in various counterfactual scenarios. The goal is to highlight some general economic forces and tradeoffs in markets with both information friction and imperfect competition.

Table 9 presents the results. The numbers are five-season discounted surpluses for the same

[37] Note that the two-season and five-season discounted returns under stick also appear to be higher than the baseline, but the difference is negligible in real terms. Any additional uncertainty, not modeled here, could discourage sellers from investing in reputation building–for example, depreciation in memory from season to season, possibility of exiting the market, etc.

average market described in Section 6.2. Details of the calculation are in Appendix E.5.

The baseline benchmark. Column 1 calculates the welfare for the baseline scenario with no quality differentiation. Using column 1 as the benchmark, I next examine the counterfactual outcomes without information friction. That is, for any quality that a seller chooses, she could immediately convey that information to consumers.

Symmetric information: one seller deviation. Column 2 considers a single seller deviation. I first solve for the seller's optimal quality and price premium for the premium pile, holding the other sellers' strategies the same as in column 1. The optimal quality of the premium pile is 0.769, much higher than that of the normal pile. The seller's net profit is almost 7 times higher than baseline. This result demonstrates that without information friction, baseline cannot be an equilibrium as there is a large single-seller profitable deviation.

Symmetric information: separating equilibrium. Column 3 computes the equilibrium outcome under symmetric information. For each γ_H and m_H chosen by the other sellers, I first solve for the optimal γ_H^* and m_H^* of the sample seller. A symmetric Nash equilibrium is found by searching for the fixed point. Here and in subsequent analyses, I focus on the best equilibrium for sellers in case of multiple equilibria. We see that competition puts a downward pressure on price and increases quality. Consumer surplus is significantly higher than that in column 2 because of the lowered price and enlarged choice set. A comparison of the total surplus in columns 1 and 3 shows that information friction result in a welfare loss of about 66.4% in this setting.

Symmetric information: first best. Column 4 solves for the first-best outcome. The key takeaway is that in this setting the welfare loss caused by market power (column 3 versus column 4) is small relative to that caused by the information problem (column 1 versus column 3), suggesting that these markets are already quite competitive. Next, I turn to welfare under asymmetric information.

Asymmetric information: one seller. The bottom panels of columns 5 and 6 compute the discounted sum of surpluses, taking into account the learning process. Compared to column 1, the increases in total surplus are 49k and 65k RMB for the non-incentive and incentive cases respectively. Balk of the gains comes from gain in consumer surplus as a result of both enlarged choice set and allocative efficiency (i.e., allowing high-valuation consumers self-sort into buying higher quality, albeit more expensive, product). In fact, the total gains are on par with the cost of a laser machine. While an individual seller would not undertake such an investment, a third-party could invest in the technology and subsidize/rent it to the sellers. The result also implies a profitable entry opportunity of a large upstream firm to invest in the technology and build a reputation for quality over time.

Asymmetric information: competition. Column 7 computes the symmetric Nash equilibrium when all sellers in a market are given access to the new technology and simultaneously choose once-for-all quality and price premium.[38] This exercise assumes that initial beliefs and learning for the premium option are the same under the case when it is provided by all sellers and by a single seller. We see that competition induces sellers to provide higher quality (compare to the monopoly case in column 5); however, quality is still quite low compared to the first-best (column 4). This is because competition on price may in fact discourage quality.

To further highlight this tradeoff, imagine a counterfactual policy in which government could regulate the price for the premium product and still let sellers compete on the quality dimension. We can think of this as analogous to the first-best but under asymmetric information when it is hard to directly enforce quality. The result is shown in the last column of Table 9. In line with the discussion above, the social planner would want to set a higher price premium to ease competition, which leads to higher quality provision compared to the competition case in column 7. That being said, the additional welfare gain is small because higher price also directly discourages sales, and beliefs take an even longer time to take off.

In reality, when multiple sellers introduce the same new experience goods, learning dynamics can be quite different from the monopoly case. For example, learning may be correlated across sellers, in which case one seller's bad behavior can adversely affect the others. Such reputational spillovers may further discourage quality provision. The current study is limited in investigating the market dynamics under such scenarios as only one seller was treated in each market; the other sellers did not strategically respond because they did not have access to the new technology. A possible extension is to vary the number of sellers treated in a market. I leave that as a potential avenue for exploration in the future.

[38] There is another low-quality equilibrium with $\gamma_H^* = 0.4$ and $m^* = 0.12$. See Appendix Table 14.

8 Conclusion

This study empirically examines the dynamic interaction between sellers and consumers in an experience good market setting. I find that information frictions and a fragmented market lead to significant under-provision of quality. Though there is a high demand for quality, trust cannot be established under the existing "contaminated" signaling technology. While there is a new technology that could enhance consumer learning and facilitate trust building, small individual sellers do not have the incentive to invest in this technology. The results suggest that third-party interventions that subsidize the initial demand and learning process could enhance welfare. The results also indicate a profitable entry opportunity for a large upstream firm.[39]

Though the exact learning processes and quality provision technologies are different for different products, the study highlights three broad takeaways:

First, a good reputation takes time to establish, as is the case with the Whole Foods brand in the United States. Such institutions may eventually emerge as consumers get richer and demand higher quality. In developing countries that currently lack such reputable entities, present market beliefs and learning dynamics matter for firms' incentive to provide quality. Markets can get stuck in a low-trust-low-quality equilibrium. In such an environment, it can be hard for a single firm to signal its quality and establish trust; hence, firms' incentive to provide quality is also low, which breeds more mistrust tomorrow. This discussion highlights an important externality due to collective reputation (Tirole, 1996).

Second, like these local retail markets, many industries in developing countries are characterized by fragmented markets with a large number of small players. While competition helps to expand sales, it can also discourage quality improvements since small firms may not find it profitable to undertake innovation activities that require large fixed costs. Competition may further dissipate the returns of investing in quality.

Finally, while the market-based reputation mechanism offers alternative solutions to address the information problem besides direct government quality control, it could break down without a strong regulatory and legal environment. Pessimistic beliefs under the stickers are partly due to past counterfeiting activities. Therefore, effective anti-counterfeiting campaigns and brand protection measures are crucial in restoring trust among consumers and strengthening firms'

[39] Indeed, a large Chinese agricultural company, Hebei Shuangxing Seed Co., Ltd., is starting a new business venture to contract with farmers, invest in high quality production of watermelons and establish a premium brand using the new technology.

incentive of inventing and introducing new high-quality products.

References

AAFC (2014): "Consumer and Retail Trends in China," Discussion paper, Agriculture and Agri-Food Canada.

Allen, T. (2014): "Information frictions in trade," Econometrica, 82(6), 2041–2083.

Atkin, D., A. K. Khandelwal, and A. Osman (2017): "Exporting and firm performance: Evidence from a randomized experiment," The Quarterly Journal of Economics, 132(2), 551–615.

Bagwell, K. (2007): "The economic analysis of advertising," Handbook of industrial organization, 3, 1701–1844.

Banerjee, A. V. (2013): "Microcredit under the microscope: what have we learned in the past two decades, and what do we need to know?," Annu. Rev. Econ., 5(1), 487–519.

Banerjee, A. V., and E. Duflo (2000): "Reputation Effects And The Limits Of Contracting: A Study Of The Indian Software Industry," The Quarterly Journal of Economics, 115(3), 989–1017.

Bar-Isaac, H., and S. Tadelis (2008): Seller reputation. Now Publishers Inc.

Bardhan, P., D. Mookherjee, and M. Tsumagari (2013): "Middlemen Margins and Globalization," American Economic Journal: Microeconomics, 5(4), 81–119.

Björkman-Nyqvist, M., J. Svensson, and D. Yanagizawa-Drott (2013): "The market for (fake) antimalarial medicine: Evidence from uganda," .

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does Management Matter? Evidence from India*.," Quarterly Journal of Economics, 128(1).

Cabral, L., and A. Hortacsu (2010): "The dynamics of seller reputation: Evidence from ebay," The Journal of Industrial Economics, 58(1), 54–78.

Cai, J., and A. Szeidl (2017): "Interfirm relationships and business performance," The Quarterly Journal of Economics.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," The Review of Economics and Statistics, 90(3), 414–427.

Ching, A. T., T. Erdem, and M. P. Keane (2013): "Invited paper-learning models: An assessment of progress, challenges, and new developments," Marketing Science, 32(6), 913–938.

De Loecker, J., and P. K. Goldberg (2014): "Firm performance in a global market," Annu. Rev. Econ., 6(1), 201–227.

De Mel, S., D. McKenzie, and C. Woodruff (2008): "Returns to capital in microenterprises: evidence from a field experiment," The Quarterly Journal of Economics, 123(4), 1329–1372.

Dickstein, M. J. (2014): "Efficient provision of experience goods: Evidence from antidepressant choice," .

Fafchamps, M. (2002): "Spontaneous market emergence," Topics in Theoretical Economics, 2(1).

Fang, H. (2001): "Social culture and economic performance," American Economic Review, pp. 924–937.

Foster, L., J. Haltiwanger, and C. Syverson (2016): "The slow growth of new plants: Learning about demand?," Economica, 83(329), 91–129.

Grace, D., K. Roesel, and T. Lore (2014): "Food safety in informal markets in developing countries: An overview," .

Harrison, A., and A. Rodríguez-Clare (2009): "Trade, foreign investment, and industrial policy for developing countries," Handbook of Development Economics, 5, 4039–4214.

Javorcik, S. B. (2004): "Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages," The American Economic Review, 94(3), 605–627.

Jensen, R., and N. Miller (2016): "Information, Demand and the Growth of Firms: Evidence from a NAtural Experiment in India."," .

Jin, G. Z., and A. Kato (2006): "Price, quality, and reputation: Evidence from an online field experiment," The RAND Journal of Economics, 37(4), 983–1005.

Jin, G. Z., and P. Leslie (2009): "Reputational incentives for restaurant hygiene," American Economic Journal: Microeconomics, 1(1), 237–267.

Klein, B., and K. B. Leffler (1981): "The role of market forces in assuring contractual performance," The Journal of Political Economy, 89(4), 615–641.

Klein, T. J., C. Lambertz, and K. O. Stahl (2016): "Market transparency, adverse selection, and moral hazard," Journal of Political Economy, 124(6), 1677–1713.

Kranton, R. E. (2003): "Competition and the incentive to produce high quality," Economica, 70(279), 385–404.

Kugler, M., and E. Verhoogen (2012): "Prices, plant size, and product quality," The Review of Economic Studies, 79(1), 307–339.

Leggett, C. G. (2002): "Environmental valuation with imperfect information the case of the random utility model," Environmental and Resource Economics, 23(3), 343–355.

List, J. A. (2006): "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," Journal of Political Economy, 114(1), 1–37.

Macchiavello, R. (2010): "Development uncorked: Reputation acquisition in the new market for Chilean wines in the UK," .

Macchiavello, R., and A. Morjaria (2015): "The value of relationships: evidence from a supply shock to Kenyan rose exports," The American Economic Review, 105(9), 2911–2945.

Macchiavello, R., A. Morjaria, et al. (2016): "Competition and relational contracts: Evidence from rwanda's coffee mills," .

Qian, Y. (2008): "Impacts of entry by counterfeiters," Quarterly Journal of Economics, 123(4), 1577–1609.

Shapiro, C. (1982): "Consumer information, product quality, and seller reputation," The Bell Journal of Economics, 13(1), 20–35.

Startz, M. (2016): "The value of face-to-face: Search and contracting problems in Nigerian trade," .

Tirole, J. (1996): "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)," The Review of Economic Studies, 63(1), 1–22.

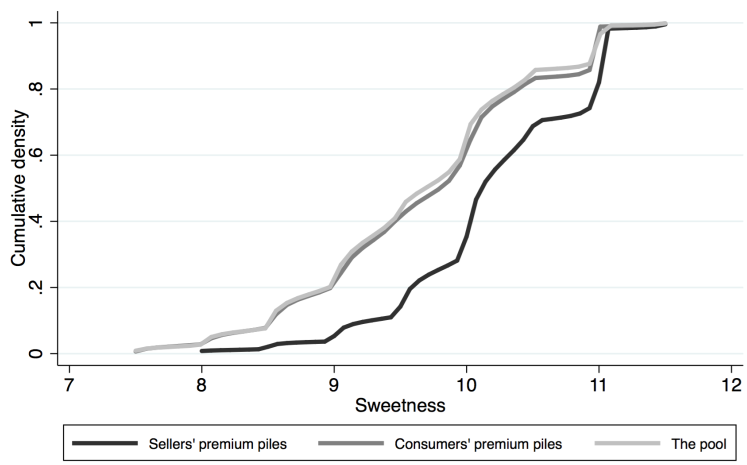
Train, K. E. (2009): Discrete choice methods with simulation. Cambridge University Press.

Verhoogen, E. A. (2008): "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," The Quarterly Journal of Economics, 123(2), 489–530.

Villas-Boas, J. M. (2004): "Consumer learning, brand loyalty, and competition," Marketing Science, 23(1), 134–145.

WHO (2007): "Countries urged to bemore vigilant about food safety," .

Figure 1: Asymmetric Information Between Sellers and Consumers



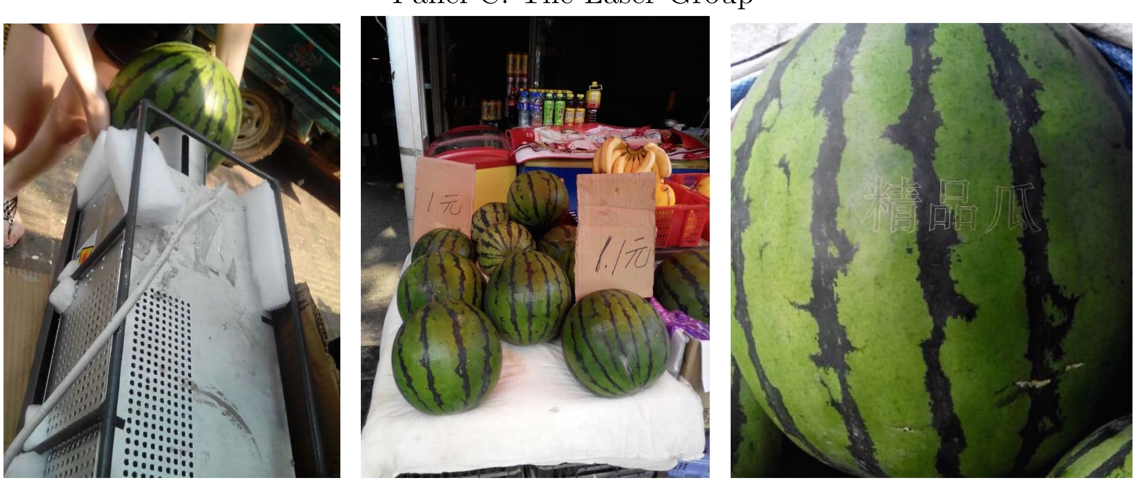
Note: This figure shows the empirical cumulative quality distribution for: (1) all 300 randomly picked watermelons used in the sorting tests; (2) the premium piles sorted by sellers; (3) the premium pile sorted by consumers. Quality is measured using a sweetness meter. For each watermelon, two measures are taken, one at the center and the other at the side, and the measures are then averaged. Details of the sorting test are described in Appendix D.1

Figure 2: Pictures of the Branding Treatments

Panel A. The Label-less Group

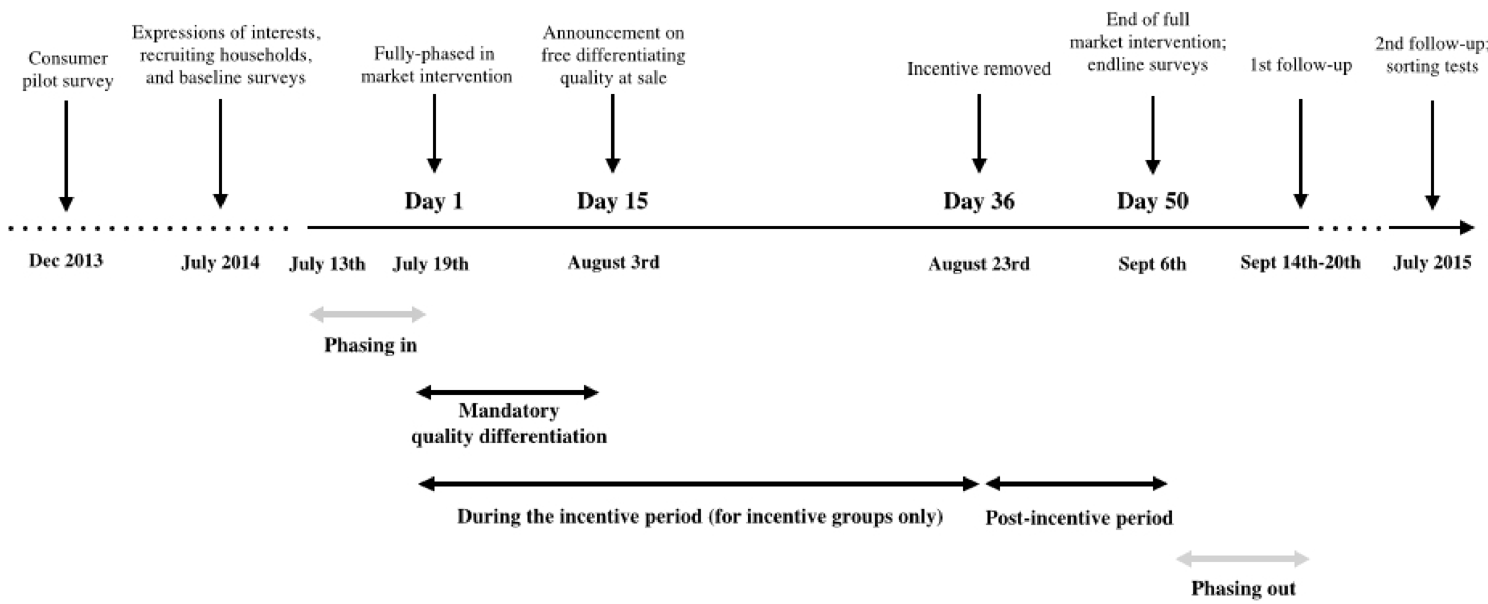
Panel B. The Sticker Group

Panel C. The Laser Group



Note: This figure depicts the actual implementation of the branding treatments. Sellers sold two piles of watermelons, a premium pile and a normal pile, and put up two price boards. Surveyors visited the markets every morning and branded the watermelons in the premium pile. Nothing was done for the label-less group (Panel A). For the sticker group, a sticker label reading "premium watermelons" was pasted on the watermelons (Panel B). For the laser group, the same words were printed on the watermelons using a laser-engraving machine (Panel C).

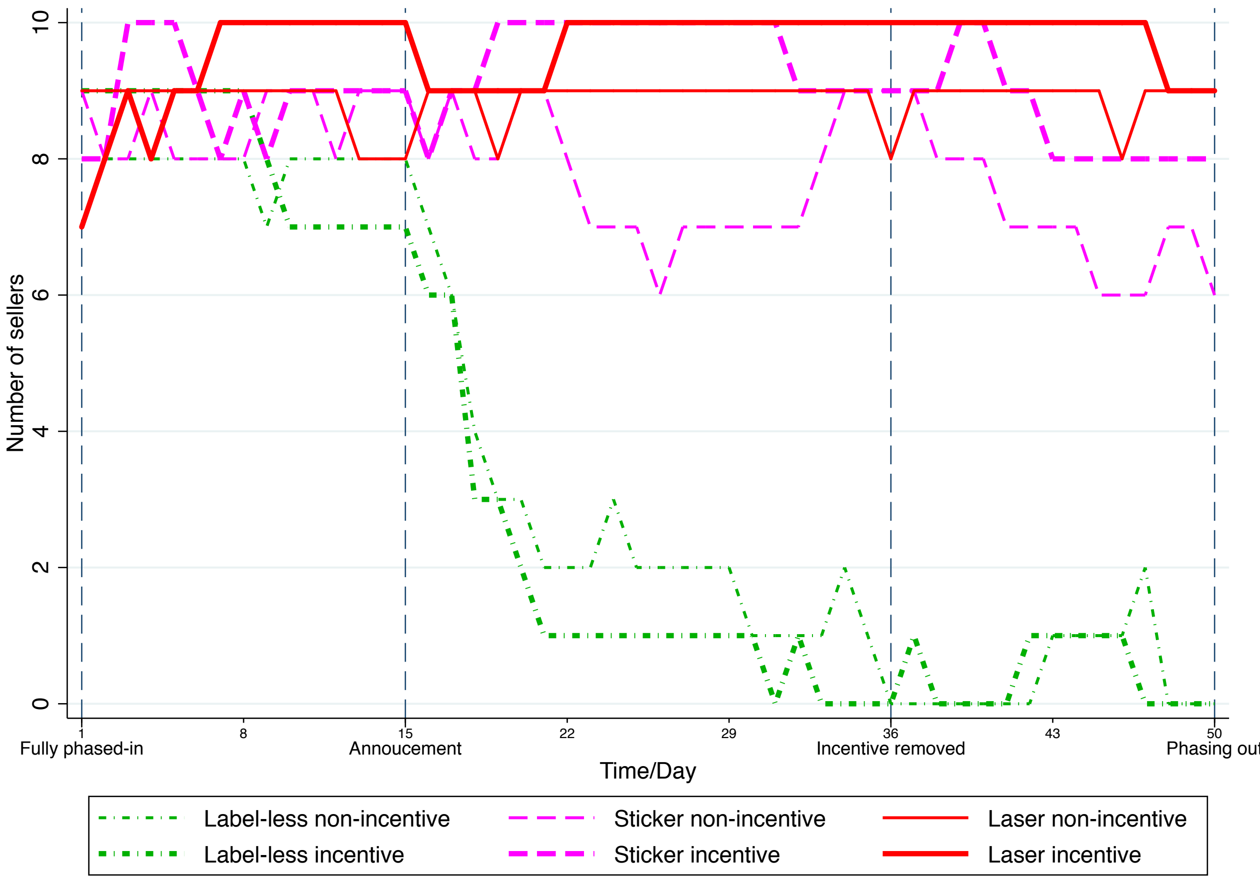
Figure 3: Timeline of the Study



Note: This figure gives an overview of the time of the study.

1. A consumer pilot survey was conducted in December 2013 to elicit consumers' perceptions of different branding technologies.
2. Expressions of interests and baseline surveys were conducted in July 2014.
3. The market intervention was rolled in from July 13 to 19, 2014. July 19 is defined to be day 1 of the full-market intervention.
4. All sellers were asked to experiment with quality differentiation for the first 2 weeks, from July 19 to August 3. (In order to participate in the experiment, sellers signed an agreement form at the beginning of the period that they would experiment with quality differentiation for the first two weeks. It was made clear to them that the research team would not interfere in any other aspect of their business, including price setting and quality choice. All sellers received a weekly compensation of 100 RMB for taking part in the study and recording daily sales data.) An announcement was made to all sellers on August 3 that they were free to differentiate or not afterwards.
5. On August 23, 35 days (6 weeks) into the intervention, the incentive (for the incentive groups) was lifted.
6. September 6 is the last day of the full-market intervention. An endline survey was conducted at surveyors' final visits to sellers' stores. Most of data analysis focuses on the period from July 19 (day 1) to September 6 (day 50).
7. The market intervention was gradually phased out from September 6 to September 12, 2014.
8. A short follow-up survey was conducted from September 14 to 20, 2014, and another one was conducted a year later, in July 2015.

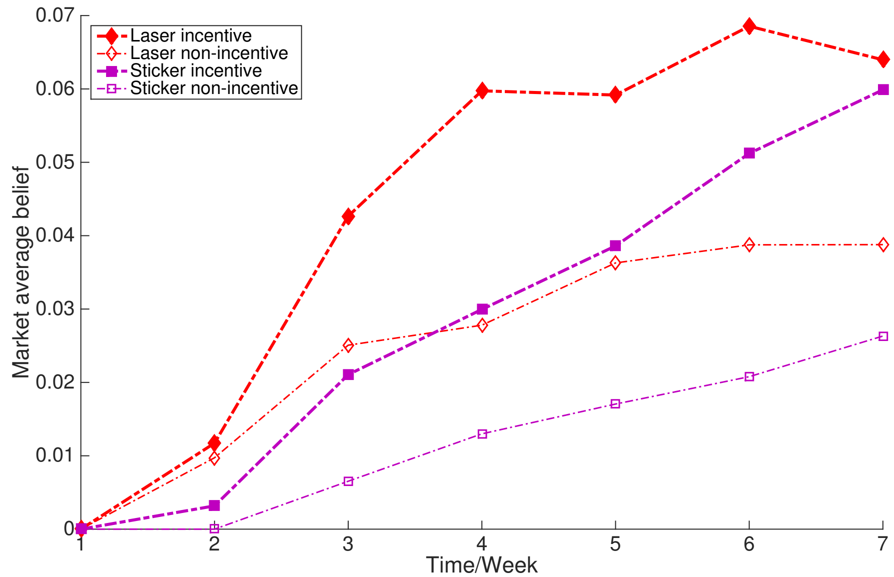
Figure 4: Quality Differentiation at Sale



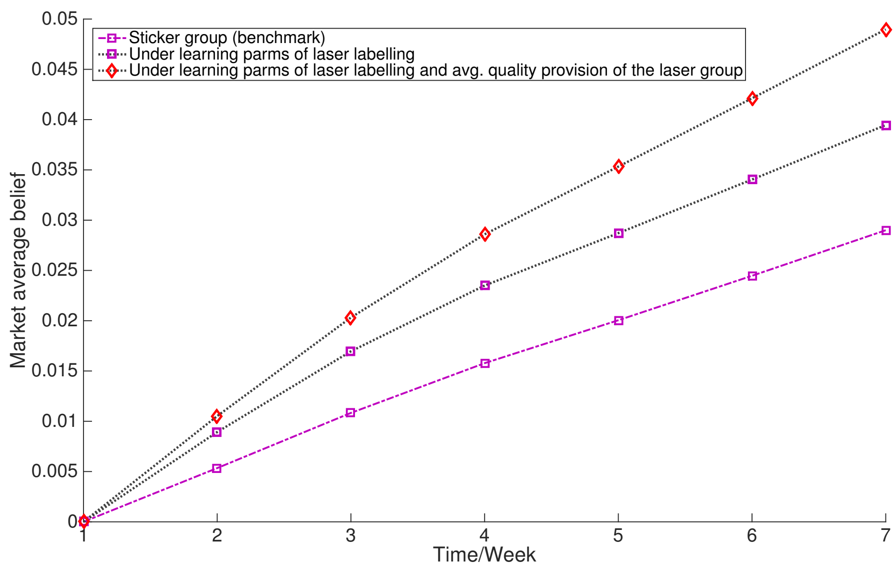
Note: This figure plots the number of sellers who differentiated quality at sale in each treatment group over time. The time axis runs from July 19 (day 1) to September 6 (day 50), 2014, corresponding to the period of the fully phased-in market intervention. The panel is not balanced because not all sellers operated their businesses on all days. Though all sellers signed an agreement at baseline that they would experiment with quality differentiation for the first two weeks, two sellers from the label-less group reneged from the beginning.

Figure 5: Beliefs Evolution

Panel A. Average Beliefs Evolution by Treatment Group

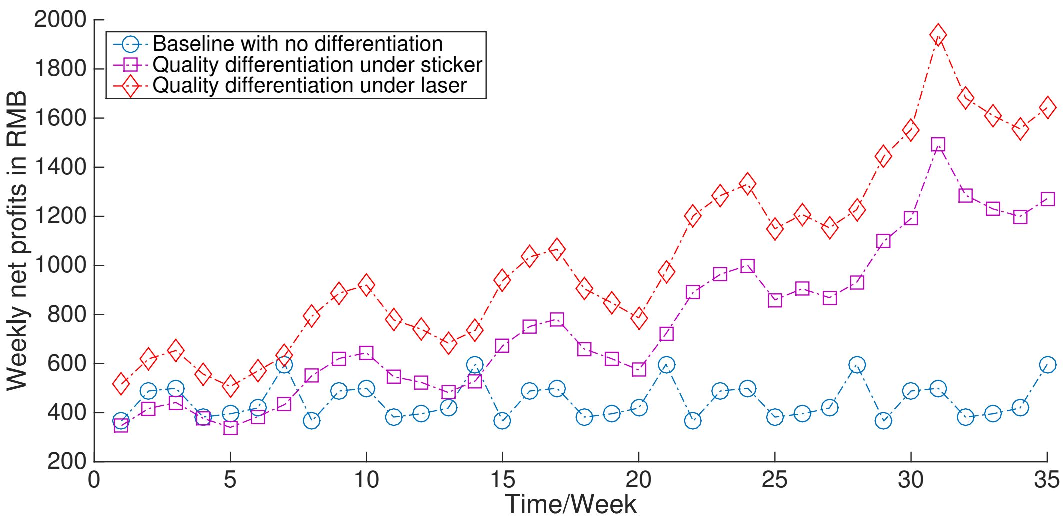


Panel B. Counterfactual Beliefs Evolution



Note: This figure plots the average beliefs evolution about the quality of the premium pile. Panel A plots the market average beliefs calculated using the estimated prior beliefs (see Table 7) and the actual experience realizations for households in each treatment group. In particular, I take the demand estimates in column 1 of Table 7 and feed them through the actual purchasing and experience realizations to compute the posterior for each household in each period. I then average that across all households in a given treatment group to get the group average beliefs. Panel B simulates the counterfactual beliefs evolution for the sample of households in the sticker group under three different scenarios: (1) under sticker group's average empirical quality (measured in terms of the empirical satisfaction rate for sticker-labeled watermelons); (2) the same quality as in (1) but replacing the prior beliefs with that under laser; (3) replacing both the prior beliefs and the average empirical quality with that for the laser group. The simulation procedure is discussed in Appendix E.4.

Figure 6: Net Profits Evolution



Note: This figure plots the simulated net profits evolution (sales profits minus effort costs) for a seller facing the hypothetical *average market* under the following three scenarios: (1) baseline with no differentiation; (2) quality differentiation under laser branding and the average empirical policies (price and quality) of the laser non-incentive group; (3) quality differentiation under sticker branding but following the same policies as (2). Details for constructing the hypothetical market is explained in Section 6.2 and Appendix E.3. The simulation procedure is discussed in Appendix E.4.

Table 1: Baseline Summary Statistics

	Observations	Median	Mean	Std. Dev
Panel A. Community and market characteristics				
Size measured in the number of housing units	60	1350	1915	1930
Housing price (in thousand RMB/meter^2)	60	8.95	8.291	1.594
Fraction of elderly	60	0.25	0.28	0.123
Distance to the nearest supermarket (in kilometer)	60	1.5	1.567	1.046
Years since establishment	60	15.5	17.633	11.242
Number of competitors in the local market	60	3	3.533	2.273
Panel B. Seller characteristics				
Gender (female=1 and male=0)	60	0	0.483	0.504
Age	60	42	41.067	9.189
Years of schooling	59	9	10.254	2.509
Selling fruits as primary income source (dummy)	60	1	0.95	0.22
Selling fruits only in the summer (dummy)	60	0	0.033	0.181
Planning to stop selling fruits (dummy)	60	0	0.017	0.129
Number of years selling fruits	60	8	9.017	6.035
Number of years selling fruits at this location	60	6.5	7.867	6.239
Planning to relocate (dummy)	60	0	0	0
Purchasing from fixed wholesaler(s) (dummy)	60	0	0.217	0.415
Panel C. Household characteristics				
Household size	658	3.5	3.76	1.366
Fraction of elderly	657	0	0.169	0.272
Fraction of female	657	0.5	0.498	0.154
Household monthly income (in thousand RMB)	647	4	5.250	3.235
Fruit as % of total food consumption	602	30	32.01	17.906
Watermelon as % of total fruit consumption	626	30	35.627	25.292
Number of watermelons consumed per week	654	1	1.308	.695
Local markets as main purchase source (dummy)	675	1	0.756	0.43
Supermarkets as main purchase source (dummy)	675	0	0.227	0.419
Willingness to pay for quality (RMB/Jin)	633	2	1.926	0.312

Note: This table provides the summary statistics for sample characteristics of communities, sellers and households measured in the baseline surveys. In total, 60 sellers in 60 communities (markets) and 675 households were recruited for this study. Variation in the number of observations are due to missing responses in the baseline surveys. To elicit willingness to pay for quality, households were asked in the baseline survey to consider a hypothetical situation wherein two piles of watermelons are sold in the local markets: one pile of ordinary quality sells at 1.5 RMB/Jin; the other of premium quality sells at a higher price. Surveyors announced the premium price from high to low and recorded the highest number that led to the choice of the premium pile. Prices (in RMB/Jin) were announced in the following order: 2.5, 2.2, 2, 1.9, 1.8, 1.7, 1.6, and 1.5.

Table 2: Purchasing Dynamics under Different Branding Technologies

	Households in the Laser Markets		Households in the Sticker Markets	
	(1)	(2)	(3)	(4)
Panel A. Purchasing decision of the premium pile				
Lagged avg. satisfaction rating	0.280**		0.049	
	(0.090)		(0.044)	
Lagged % of very good experiences		0.454**		0.110
		(0.129)		(0.075)
Observations	165	167	183	183
Panel B. Purchasing decision of the normal pile				
Lagged avg. satisfaction rating	0.035		-0.014	
	(0.029)		(0.039)	
Lagged % of very good experiences		0.010		-0.016
		(0.032)		(0.086)
Observations	520	576	497	530
Household Baseline Controls	✓	✓	✓	✓
Week Fixed Effects	✓	✓	✓	✓

Note: This table examines the purchasing dynamics under different branding technologies. Each observation is at the household-week level. The dependent variable for Panel A is whether the household has purchased any watermelon from the premium pile for a given week. The dependent variable for Panel B is the corresponding purchasing dummy for the normal pile. The purchasing dummies are regressed on two measures of lagged experiences: (1) the average lagged satisfaction rating (ranging from 1 to 5) of all premium watermelons purchased prior to the period; (2) the percentage of past consumption experiences that attained the highest satisfaction rating of 5. Note that if a household has never purchased any premium watermelons, these lagged experience measures are not defined. Therefore, the coefficients are only estimated from household-week observations for which a positive number of premium watermelons have been consumed by the household prior to the given week. All regressions control for week fixed effects and the following set of household baseline characteristics: household size, percentage of elderly, monthly income, average number of watermelons consumed per week reported in the baseline survey, and the baseline self-reported willingness to pay for quality (in RMB/Jin). Standard errors are clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 3: Quality Provision by Treatment Group

Dep var: Quality of the premium pile (measured in sweetness)

	A. By branding treatments (sticker and laser)				B. By incentive treatment (during incentive)			
	Non-incentive		Incentive		Laser		Sticker	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
laser	0.711***	0.619**	0.282*	0.309**				
	(0.222)	(0.266)	(0.136)	(0.128)				
incentive					0.496*	0.563**	1.033***	1.006***
					(0.246)	(0.266)	(0.176)	(0.176)
Observations	238	238	230	230	197	197	194	194
Baseline Controls		✓		✓		✓		✓
Time Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Small sample robustness								
Permutation test (p-value):	0.008	0.026	0.077	0.076	0.067	0.081	0.000	0.001
Clustered bootstrap (p-value):	0.001	0.085	0.043	0.052	0.044	0.08	0.000	0.000
Omitted group mean	9.738		10.654		10.451		9.738	
Std. dev	(1.104)		(0.886)		(1.04)		(1.104)	

Note: This table examines quality provision by treatment group. Quality is measured in sweetness. Each observation is at the seller-check level. The key explanatory variables are the group dummies. The mean and standard deviation for the omitted group are shown in the bottom two rows. Panel A examines the heterogeneity across different branding groups. Columns 1 and 2 restrict the sample to the non-incentive groups. Columns 3 and 4 restrict to the incentive groups. Panel B examines the heterogeneity across the incentive treatment. Since sellers in the label-less group reverted back to non-differentiation after the first two weeks, the sample for this analysis includes only sellers in the sticker and laser groups. The time period is from week 1 to week 6, before the incentive was lifted. Columns 5 and 6 look within the laser group. Columns 7 and 8 look within the sticker group. All regressions control for check fixed effects. The even columns control for additional seller and community baseline characteristics: number of competitors in the local market, average housing price, and distance to the nearest supermarket. Standard errors are clustered at the seller level. Small sample robustness implements two different procedures to address the concern of a relatively small sample size. Permutation test reports the p-values for testing the null hypothesis that laser (incentive) has no effect by randomly permuting the values of laser (incentive) 1000 times while respecting seller clusters. Clustered bootstrap performs nonparametric bootstrap estimation of the regression coefficients (1000 replications). ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 4: Quality Differentiation Behavior

Sample: sticker and laser non-incentive groups

	Dep var: Quality measured in sweetness			
	A. Level		B. Diff. from the avg. pool	
	Laser	Sticker	Laser	Sticker
	(1)	(2)	(3)	(4)
Premium pile	0.735***	0.378**	0.786***	0.453**
	(0.157)	(0.163)	(0.129)	(0.172)
Observations	212	184	142	116
Seller Fixed Effects	✓	✓	✓	✓
Time Fixed Effects	✓	✓	✓	✓
Normal pile mean	9.787	9.366	0.102	-0.285
Std. dev.	(0.99)	(0.923)	(0.774)	(0.965)

Note: This table examines the quality differentiation behavior of sellers in the sticker and laser non-incentive groups. Quality is measured in sweetness. Each observation is at the seller-pile-check level. The key explanatory variable is a dummy for the premium pile. The mean and standard deviation for the normal pile are shown in the bottom two rows. The dependent variable for Panel A is in the level of the measured sweetness and that for Panel B is the difference from the market average quality. The average is computed as the average sweetness of randomly picked watermelons from the undifferentiated piles of the label-less group at each check (from week 3 and onwards). All regressions control for seller and time fixed effects. Standard errors are clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 5: Effects of the Branding Treatments on Price, Quantity and Profits

Sample: non-incentive groups

	Ln(Sales Profits)		Premium Price Δ		Premium Quantity		Normal Price Δ		Normal Quantity		Total Quantity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
sticker	0.031	-0.038	0.039**	0.045***	49.852*	49.454*	0.001	-0.001	-40.374	-55.550**	9.478	-6.096
	(0.199)	(0.196)	(0.015)	(0.015)	(28.758)	(28.506)	(0.010)	(0.009)	(24.860)	(23.831)	(39.378)	(41.676)
laser	0.297*	0.396**	0.069***	0.065***	62.041***	70.450***	-0.006	-0.001	-12.445	-4.449	49.596	66.002**
	(0.154)	(0.156)	(0.020)	(0.019)	(22.073)	(23.296)	(0.010)	(0.010)	(26.705)	(18.699)	(36.728)	(31.906)
Observations	1452	1452	1456	1456	1462	1462	1456	1456	1462	1462	1462	1462
Baseline Controls		✓		✓		✓		✓		✓		✓
Time Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Small sample robustness												
Permutation test (p-value):												
sticker	0.881	0.859	0.096	0.053	0.080	0.130	0.906	0.945	0.150	0.039	0.809	0.882
laser	0.132	0.068	0.000	0.007	0.0290	0.027	0.514	0.891	0.689	0.862	0.210	0.112
Clustered bootstrap (p-value):												
sticker	0.876	0.860	0.016	0.012	0.080	0.120	0.901	0.952	0.113	0.035	0.804	0.894
laser	0.061	0.026	0.001	0.003	0.006	0.010	0.528	0.895	0.659	0.835	0.188	0.078
Label-less Mean	4.284		0.055		56.313		0.011		180.475		236.788	
Std. dev.	(0.687)		(0.091)		(136.508)		(0.084)		(124.07)		(156.597)	

Note: This table examines sales profits, price and quantity for sellers in the non-incentive groups. Each observation is at the seller-day level. Sticker and laser are group dummies, and the omitted group is the label-less group, the mean and standard deviation for which are shown in the last two rows. Price Δ is defined to be the difference between the unit price (RMB/Jin) charged by the seller and the market average retail price. Quantity is measured in Jin and profits are measured in RMB. If a seller stops to differentiate quality at sale, the unit price of the premium pile is defined to be the same as that of the normal pile, and the sales quantity of the premium pile is coded as 0. The even columns control for the following set of seller and community baseline characteristics: number of competitors in the local market, average housing price, and distance to the nearest supermarket. All regressions control for day fixed effects. Standard errors are clustered at the seller level. Small sample robustness implements two different procedures to address the concern of a relatively small sample size. Permutation test reports the p-values for testing the null hypothesis that laser (incentive) has no effect by randomly permuting the values of laser (incentive) 1000 times while respecting seller clusters. Clustered bootstrap performs nonparametric bootstrap estimation of the regression coefficients (1000 replications). ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 6: Effects of Removing the Incentive on Quality Provision

Dep var: Quality of the premium pile (measured in sweetness)

	Laser		Sticker	
	(1)	(2)	(3)	(4)
Incentive	0.502**	0.550**	1.026***	1.034***
	(0.239)	(0.256)	(0.171)	(0.169)
Post	0.013	0.014	0.224	0.226
	(0.299)	(0.301)	(0.255)	(0.256)
Post X Incentive	-0.008	-0.008	-0.683*	-0.674*
	(0.401)	(0.405)	(0.376)	(0.380)
Observations	236	236	232	232
Seller (Market) Baseline Controls		✓		✓

Note: This table runs a difference-in-difference regression to examine the effect of removing the incentive. The dependent variable is the measured sweetness of watermelons in the premium pile. Incentive is a dummy for the incentive group. Post is a dummy for the period after the incentive was lifted (i.e. week 7 and 8). The key explanatory variable is the interaction term. Each observation is at the seller-check level. Columns 1 and 2 look within the laser groups; columns 3 and 4 look within the sticker groups. In addition, the even columns control for a set of baseline characteristics, including the number of competitors in the local market, the average housing price, and distance to the nearest supermarket. Standard errors are clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 7: Simulated Maximum Likelihood Estimation Results of Consumer Learning Models

Parameters	Baseline Model (1)		Direct Utility of Laser (2)		Correlated Learning (3)		Information Diffusion (4)	
$a_0(s)$	0.000	(-)	0.000	(-)	0.000	(-)	0.000	(-)
$b_0(s)$	2.578	(0.733)	2.383	(0.683)	2.639	(0.818)	2.453	(0.757)
$a_0(l)$	0.000	(-)	0.000	(-)	0.000	(-)	0.000	(-)
$b_0(l)$	0.938	(0.471)	1.037	(0.510)	0.995	(0.554)	0.850	(0.498)
q	0.307	(0.088)	0.313	(0.089)	0.283	(0.089)	0.309	(0.098)
θ_0	8.549	(1.197)	8.500	(1.185)	9.149	(1.577)	8.518	(1.533)
θ_1	0.346	(0.285)	0.309	(0.277)	0.373	(0.312)	0.330	(0.286)
α_0	0.169	(0.046)	0.170	(0.045)	0.166	(0.046)	0.168	(0.046)
α_1	-0.007	(0.006)	-0.007	(0.006)	-0.007	(0.006)	-0.007	(0.006)
β	0.061	(0.035)	0.062	(0.035)	0.057	(0.035)	0.057	(0.035)
$m(\eta)$	0.479	(0.195)	0.406	(0.236)	0.451	(0.108)	0.442	(0.216)
$\sigma(\eta)$	0.426	(0.182)	0.436	(0.196)	0.433	(0.188)	0.433	(0.191)
$m(\xi)$	-1.583	(0.046)	-1.585	(0.046)	-1.583	(0.046)	-1.584	(0.046)
$\sigma(\xi)$	0.784	(0.056)	0.786	(0.056)	0.784	(0.056)	0.784	(0.056)
$\Delta q(s)$	-0.081	(0.022)	-0.082	(0.023)	-0.064	(0.025)	-0.081	(0.029)
$\Delta q(l)$	-0.001	(0.012)	-0.003	(0.013)	-0.003	(0.011)	-0.003	(0.012)
$\nu(l)$	n.a.	-	0.399	(0.278)	n.a.	-	n.a.	-
$\phi_{\text{spillover}}$	n.a.	-	n.a.	-	1.218	(0.839)	n.a.	-
ϕ_{info}	n.a.	-	n.a.	-	n.a.	-	2.176	(3.597)
Market FE (abbreviated)	✓		✓		✓		✓	
Time FE (abbreviated)	✓		✓		✓		✓	
Log likelihood	-3709.749		-3708.752		-3708.578		-3708.383	
D (-2×Log(likelihood ratio))			1.993		2.341		2.732	

Note: This table shows the simulated maximum likelihood estimation results of the consumer learning models. a_0 and b_0 are constrained to be non-negative. Details for the estimation procedures are explained in Appendix E.1. Column 1 shows the estimates for the baseline model. Column 2 includes a product-specific constant ν for the premium option under laser label to account for any direct utility of laser. Column 3 incorporates correlated learning by allowing the posterior for the premium pile to enter linearly into the mean utility of the normal pile (i.e. good experiences from the premium pile may lead consumers to favor the sample seller in general). Column 4 includes a linear function of the market average beliefs (computed as the average beliefs of households in a given market at a given time) in the mean utility of the premium option to account for information diffusion. The log-likelihood ratio statistics for testing the baseline model against these alternative models are presented in the last row. Estimates for the market and time fixed effects are abbreviated from this table and are reported in Appendix Table 12. Standard errors shown in parentheses are calculated using the outer product of gradients (OPG) estimate for the asymptotic covariance matrix (see Appendix E.1 for details).

Table 8: Simulated Market Outcomes

Structural parameters
Market size : 4.5 $\times$ 194 households (to match initial sales quantity)
$\delta = 0.98$, $c = 0.64$

	Laser non-incentive		Laser incentive		Counterfactual I Prior beliefs under sticker	Counterfactual II No differentiation
	(1)		(2)		(3)	(4)
Empirical average policies						
Average quality of the undifferentiated pile ($\underline{\gamma}$)	0.300		0.300		0.300	0.300
Average quality of the premium pile ($\overline{\gamma}_H$)	0.400		0.530		0.400	0.300
Average price premium of the premium pile in RMB/Jin ($\overline{m}_H$)	0.142		0.178		0.142	0.000
Average weekly outcomes for the first season						
	Simulated	Actual	Simulated	Actual	Simulated	Simulated
Sales quantity of the premium pile (number)	53	50	58	62	41	-
Sales quantity of the normal pile (number)	81	76	80	74	48	85
Total sales quantity (number)	133	126	138	136	89	85
Total sales quantity of other sellers (number)	311	-	303	-	331	321
Sales profits (in RMB)	657	748	760	875	461	450
Net profits (sales profits minus effort costs) (in RMB)	579	-	550	-	392	450
Sales profits of other sellers (in RMB)	1,345	-	1,390	-	1,428	1,754
Simulated longer term outcomes						
Disc. Σ of net profits for two seasons (in RMB)	8,361		7,554		5,777	5,524
Disc. Σ of net profits for five seasons (in RMB)	24,408		23,165		13,281	11,367

Note: This table simulates market outcomes for the hypothetical *average market* using the estimated dynamic demand system and the estimated supply-side parameters. Details for constructing the hypothetical market are explained in Section 6.2. Column 1 simulates the market outcomes under the average empirical policies of the laser non-incentive group and column 2 does that for the laser incentive group. Column 3 performs a counterfactual exercise by replacing the learning parameters (including $a_0, b_0, \Delta q$) under laser with those under sticker (see Table 7). Column 4 simulates the outcomes for the baseline case with no quality differentiation. Details for the simulation procedures are explained in Appendix E.4.

Table 9: Welfare Effects of Information friction and Fragmented Market

	Baseline	Symmetric information			Asymmetric information			
		One seller deviation	Oligopolistic competition	First-best	One seller w/o incentive	One seller w incentive	Oligopolistic competition	Price regulation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quality and price premium								
Average quality of the undifferentiated pile ($\underline{\gamma}$)	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300
Quality of the premium pile (γ_H)	-	0.769	0.787	0.825	0.400	0.530	0.440	0.530
Price premium of the premium pile (m_H)	-	1.156	1.080	0.577	0.142	0.178	0.170	0.340
No adjustment (disc. Σ of 5 seasons)								
Sales profits	11,367	237,102	83,515	59,315	52,963	91,515	28,292	38,859
Effort costs	0	147,736	46,178	56,895	14,801	58,009	7,863	14,764
Net profits (PS_{own})	11,367	89,365	37,337	2,420	38,162	33,505	20,429	24,095
Sales profits of other sellers	44,330	23,568	335,177	241,773	31,793	21,199	102,983	149,404
Effort costs of other sellers	0	0	188,691	233,224	0	0	31,973	60,181
Net profits of other sellers (PS_{other})	44,330	23,568	146,486	8,550	31,793	21,199	71,010	89,222
Expected maximum utility in RMB (CS)	207,419	370,370	598,265	804,228	305,196	394,443	484,279	531,841
Total surplus ($= \text{PS}_{\text{own}} + \text{PS}_{\text{other}} + \text{CS}$)	263,116	483,303	782,088	815,198	375,151	449,147	575,718	645,158
Ratio relative to baseline	1.000	1.837	2.972	3.098	1.426	1.707	2.188	2.452
With adjustment (disc. Σ of 5 seasons)								
Net profits (PS_{own})	-	-	-	-	24,408	23,165	14,695	15,400
Net profits of other sellers (PS_{other})	-	-	-	-	39,357	39,134	68,011	71,448
Expected maximum utility in RMB (CS)	-	-	-	-	248,408	266,130	361,737	363,430
Total surplus ($= \text{PS}_{\text{own}} + \text{PS}_{\text{other}} + \text{CS}$)	-	-	-	-	312,173	328,429	444,443	450,278

Note: This table examines the welfare effects of information friction and market competition. The top panel solves for the optimal policies under each counterfactual scenario. Quality is the probability of being good and price premium is the difference between the prices of the premium and the normal pile, measured in RMB/Jin. The middle and bottom panel calculate the 5-season discounted sum of surpluses (in RMB) under the corresponding policies for the same hypothetical *average market* as that for Table 8 (see in Section 6.2 for details on constructing the hypothetical market). Details for calculating the consumer and producer surpluses are discussed in Section 7.

FOR ONLINE PUBLICATION

Appendix A. Appendix Tables and Figures

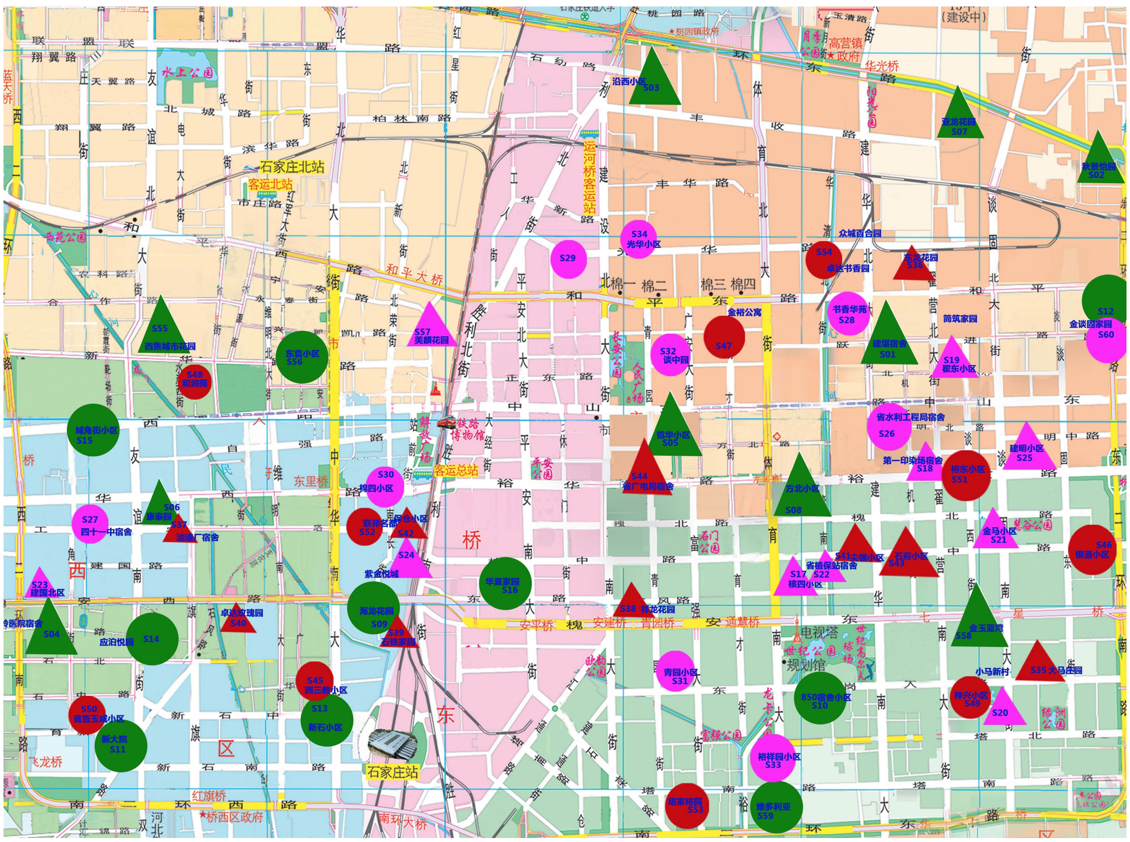
Appendix Figure 1: A Local Market

Note: This figure shows a picture of a typical local market.

Appendix Figure 2: A Sweet Meter

Note: This figure shows the picture of a sweet meter that measures the sugar content of watermelons.

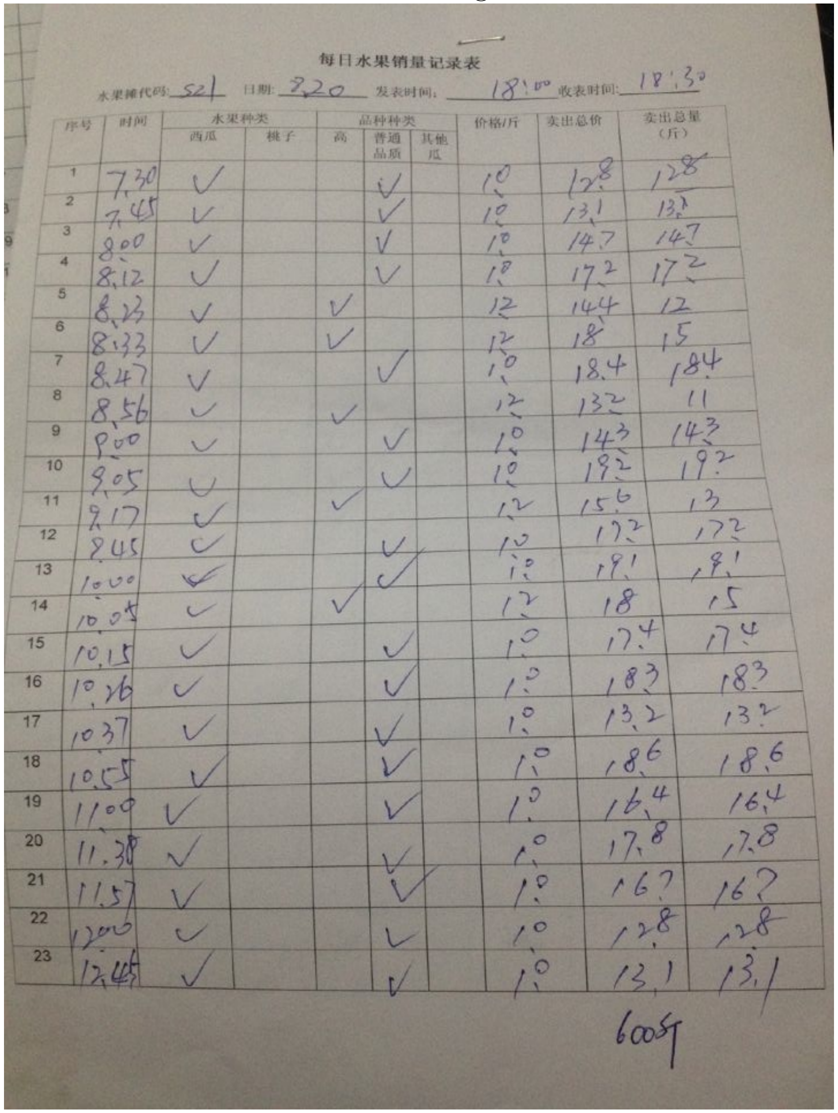
Appendix Figure 3: Map of the Randomization



Note: This figure shows the map for the urban area of Shijiazhuang (399.3 sq km) and the geographical location of the 60 sellers in the study sample. The average distance between two closest markets is about 1 km. Sellers in the label-less group are marked in green; those in the sticker group are marked in magenta; and those in the laser group are marked in red. Circle represents the incentive group and triangle represents the non-incentive group.

Appendix Figure 4: Examples of the Seller and Household Recording Sheets

A. Seller Recording Sheet



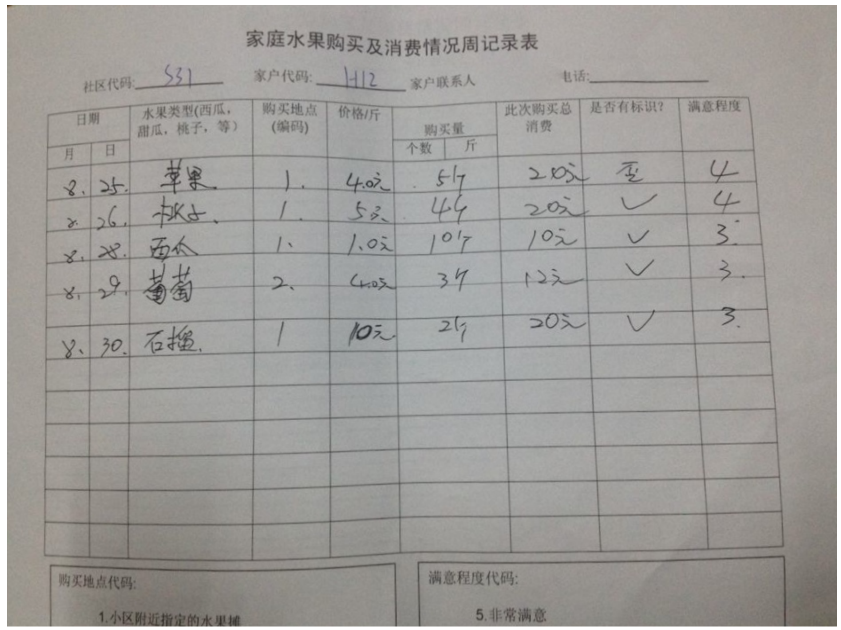
每日水果销量记录表

水果摊代码: S21　日期: 8.20　发表时间: 18:00　收表时间: 18:30

序号	时间	水果种类		品种种类			价格/斤	卖出总价	卖出总量（斤）
		西瓜	桃子	高	普通品质	其他瓜			
1	7.30	✓			✓		1.0	12.8	12.8
2	7.45	✓			✓		1.0	13.1	13.1
3	8.00	✓			✓		1.0	14.7	14.7
4	8.12	✓			✓		1.0	17.2	17.2
5	8.23	✓		✓			1.2	14.4	12
6	8.33	✓		✓			1.2	18	15
7	8.47	✓			✓		1.0	18.4	18.4
8	8.56	✓		✓			1.2	13.2	11
9	9.00	✓			✓		1.0	14.3	14.3
10	9.05	✓			✓		1.0	19.2	19.2
11	9.17	✓		✓			1.2	15.6	13
12	9.45	✓			✓		1.0	17.2	17.2
13	10.00	✓			✓		1.0	19.1	19.1
14	10.05	✓		✓			1.2	18	15
15	10.15	✓			✓		1.0	17.4	17.4
16	10.26	✓			✓		1.0	18.3	18.3
17	10.37	✓			✓		1.0	13.2	13.2
18	10.55	✓			✓		1.0	18.6	18.6
19	11.00	✓			✓		1.0	16.4	16.4
20	11.38	✓			✓		1.0	17.8	17.8
21	11.57	✓			✓		1.0	16.7	16.7
22	12.00	✓			✓		1.0	12.8	12.8
23	12.45	✓			✓		1.0	13.1	13.1

600斤

B. Household Recording Sheet



家庭水果购买及消费情况周记录表

社区代码: S31　家户代码: H12　家户联系人　电话:

日期		水果类型(西瓜，甜瓜，桃子，等)	购买地点(编码)	价格/斤	购买量		此次购买总消费	是否有标识?	满意程度
月	日				个数	斤			
8.	25.	苹果.	1.	4.0元		5斤	20元	否	4
8.	26.	桃子.	1.	5元		4斤	20元	✓	4
8.	28.	西瓜	1.	1.0元		10斤	10元	✓	3.
8.	29.	葡萄	2.	4.0元		3斤	12元	✓	3.
8.	30.	石榴.	1	10元		2斤	20元	✓	3.

购买地点代码:

1.小区附近指定的水果摊

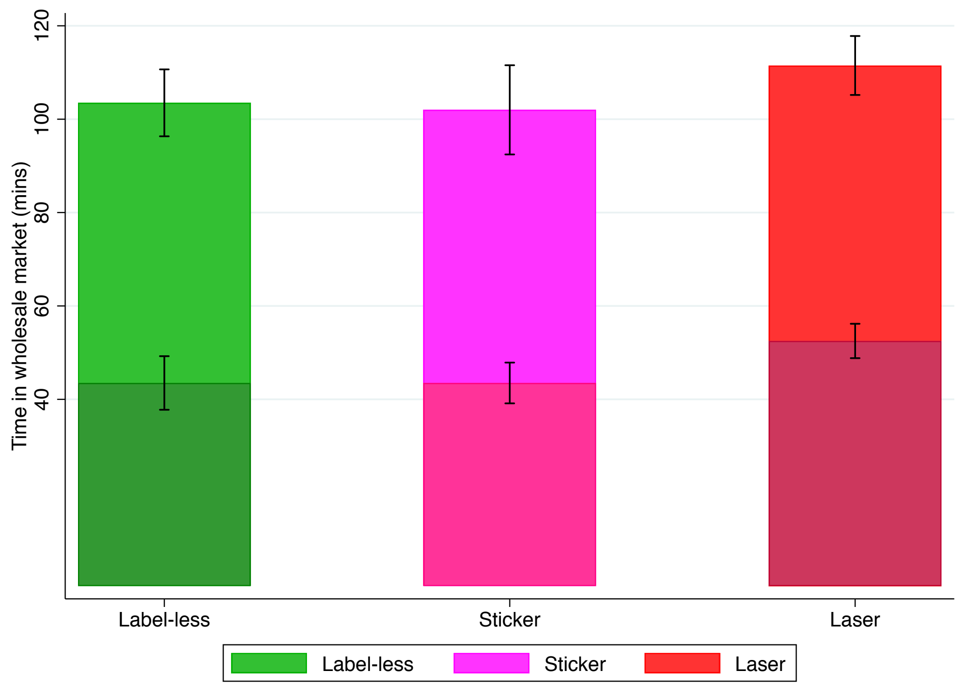
满意程度代码:

5.非常满意

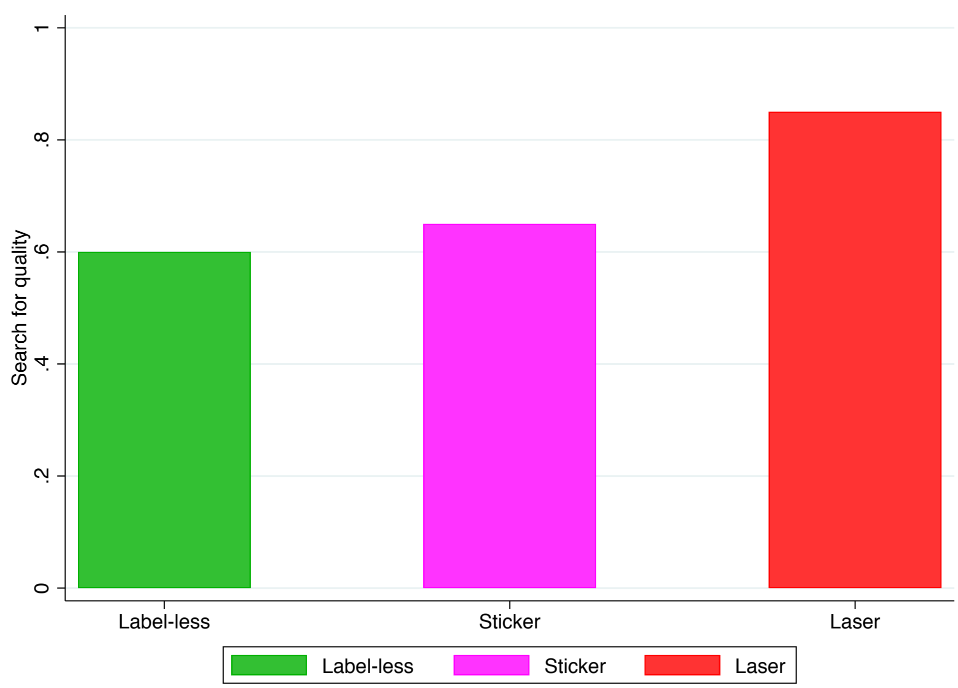
Note: This figure shows an example of a seller recording sheet (daily) and a household recording sheet (weekly). See Section 4.2.

Appendix Figure 5: Sourcing Efforts in Wholesale Markets

A. Time (Minutes) Spent in Wholesale Market

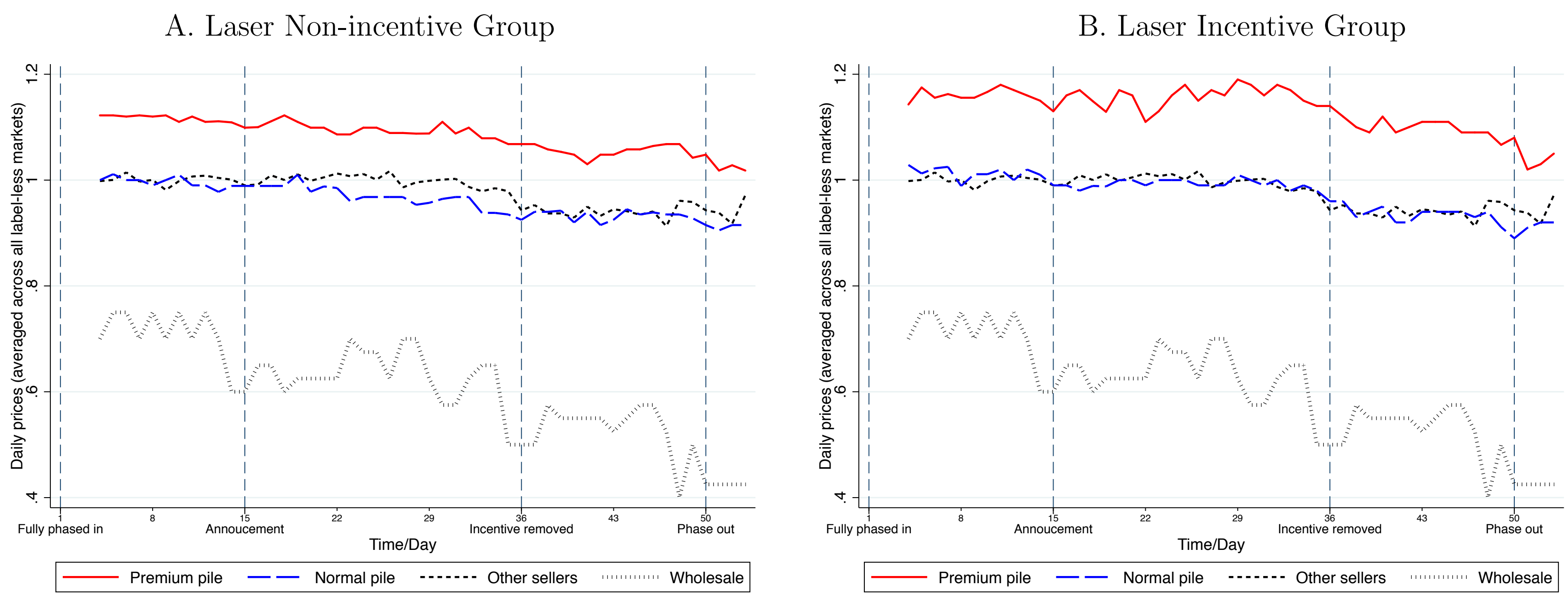


B. Search for Quality Watermelons



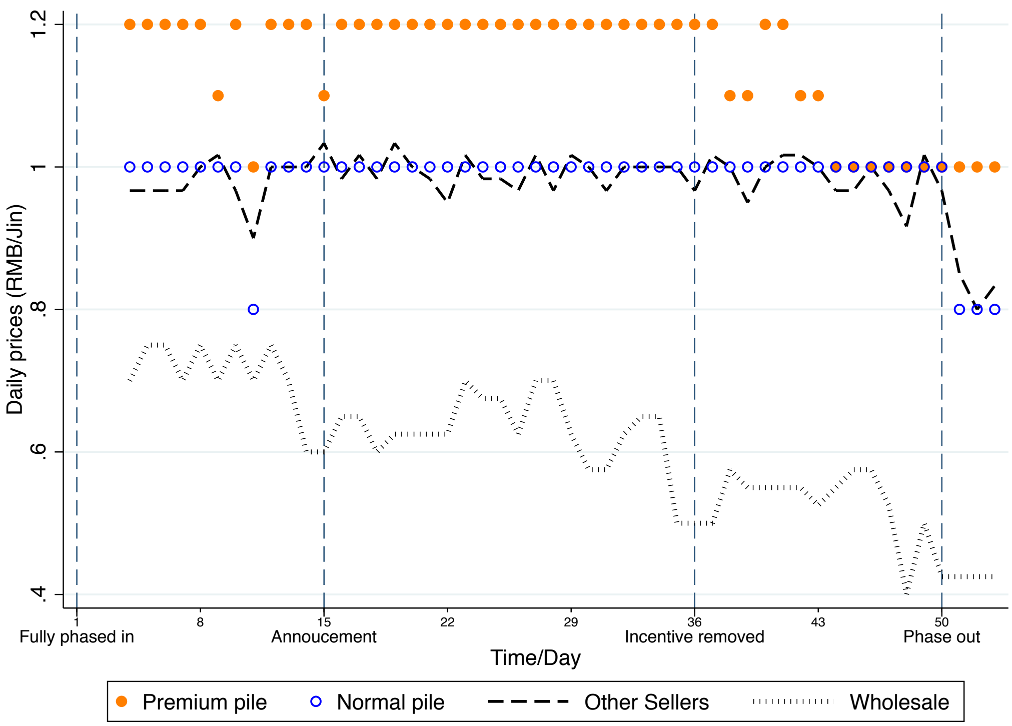
Note: This figure compares sellers' self-reported behavior in the wholesale market. Panel A plots the mean and standard deviation for total time spent in the wholesale market (lighter color) and time spent on sourcing watermelons (darker color). Panel B plots the fraction of sellers who indicated that he or she intentionally searched for quality watermelons.

Appendix Figure 6: Price Dynamics (Laser Groups)



Note: This figure plots the average daily price for the laser groups. The dotted lines plot the daily wholesale price; the short-dash lines plot the daily average local market price charged by other sellers in these markets; the long-dash lines plot the daily average price of the normal pile charged by the sample sellers; finally, the solid lines plot the daily average price of the premium pile.

Appendix Figure 7: Seller's Pricing Behavior: a Typical Market

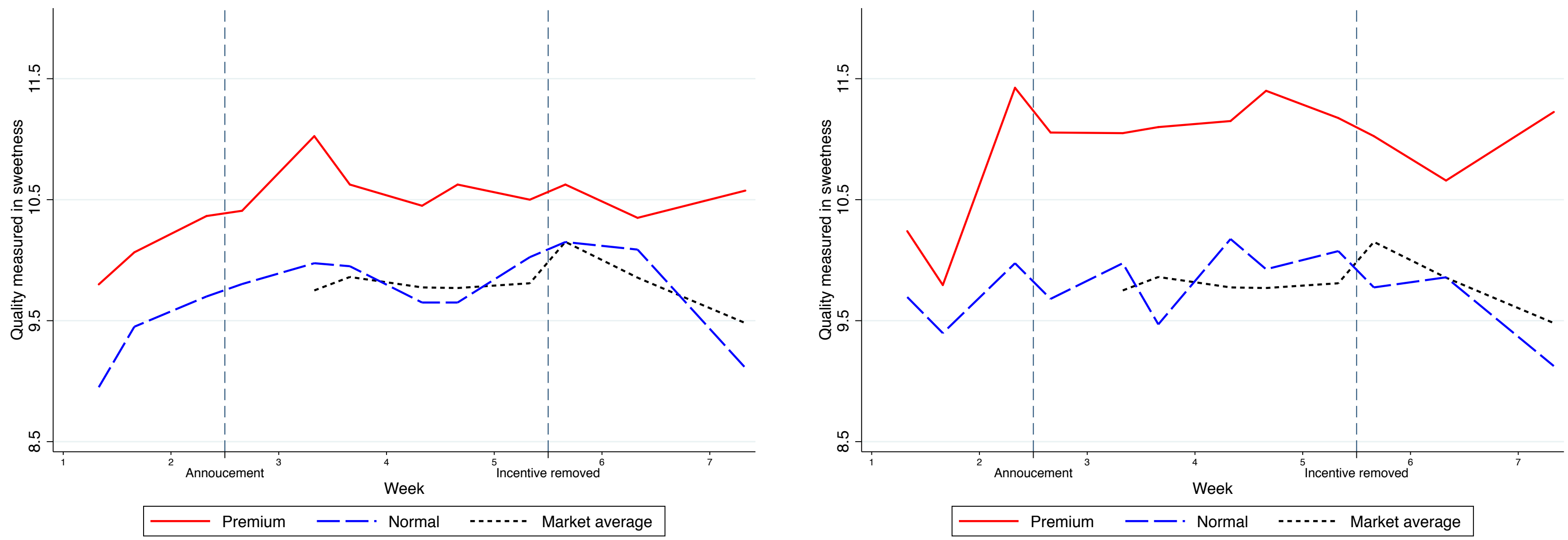


Note: This figure depicts the pricing behavior of a typical seller. The dotted line plots the daily wholesale price; the short-dash line plots the daily average local market price charged by the other sellers in the market; the empty circles mark the price of the normal pile; the solid circles mark the price of the premium pile.

Appendix Figure 8: Quality Dynamics (Laser Groups)

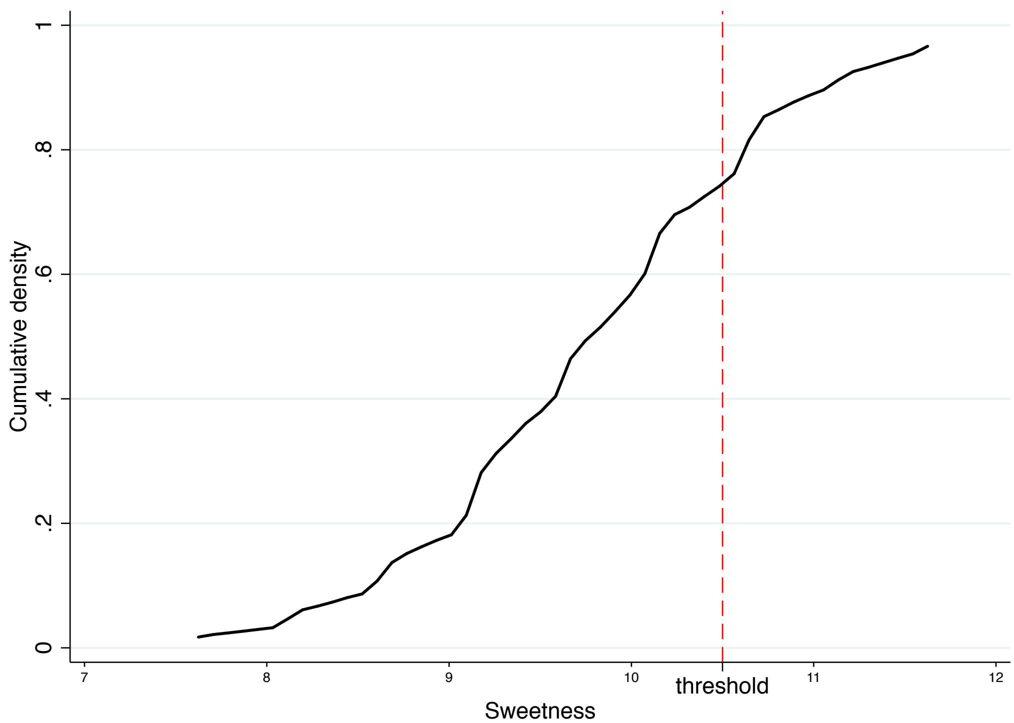
A. Laser Non-incentive Group

B. Laser Incentive Group



Note: This figure plots the average quality for the laser groups. Quality is measured in sweetness from the biweekly quality checks. The dotted lines plot the average quality of the undifferentiated pile, calculated using the sweetness of watermelons from the label-less group after sellers reverted back to non-differentiation. Since most label-less sellers only reverted back to non-differentiation after the first two weeks, this measure is only defined from week 3 and onwards. The long-dash lines plot the average quality of the normal pile and the solid lines plot the average quality of the premium pile.

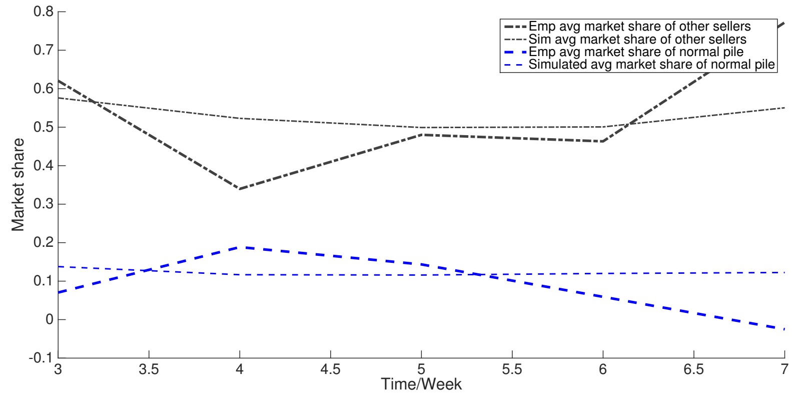
Appendix Figure 9: Distribution of Sweetness for the Undifferentiated Pile



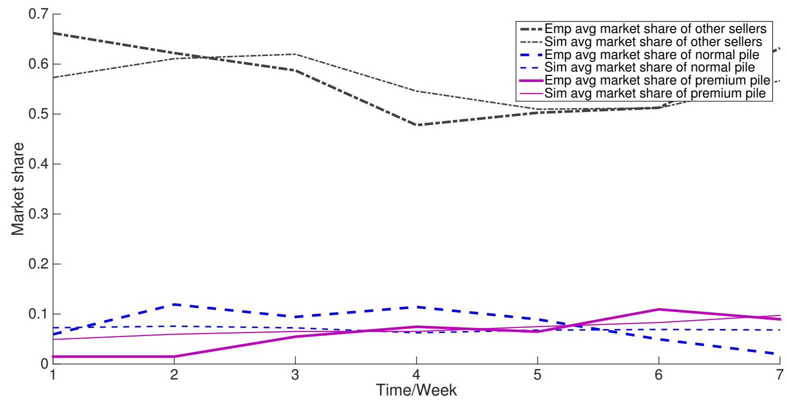
Note: This figure plots the empirical cumulative distribution of the measured sweetness for the label-less group after the sellers reverted back to non-differentiation. The threshold, 10.5, marks the criterion for receiving the incentive. The corresponding cumulative density is 0.73.

Appendix Figure 10: Model Fit: Simulated vs Actual Market Shares

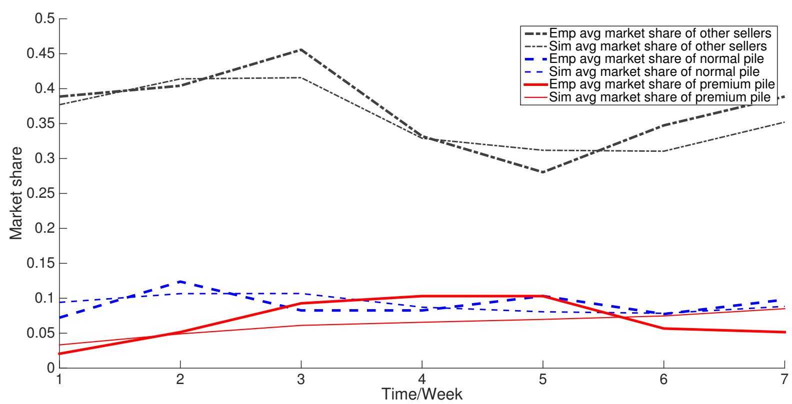
Panel A. The label-less Markets



Panel B. The Sticker Markets



Panel C. The Laser Markets

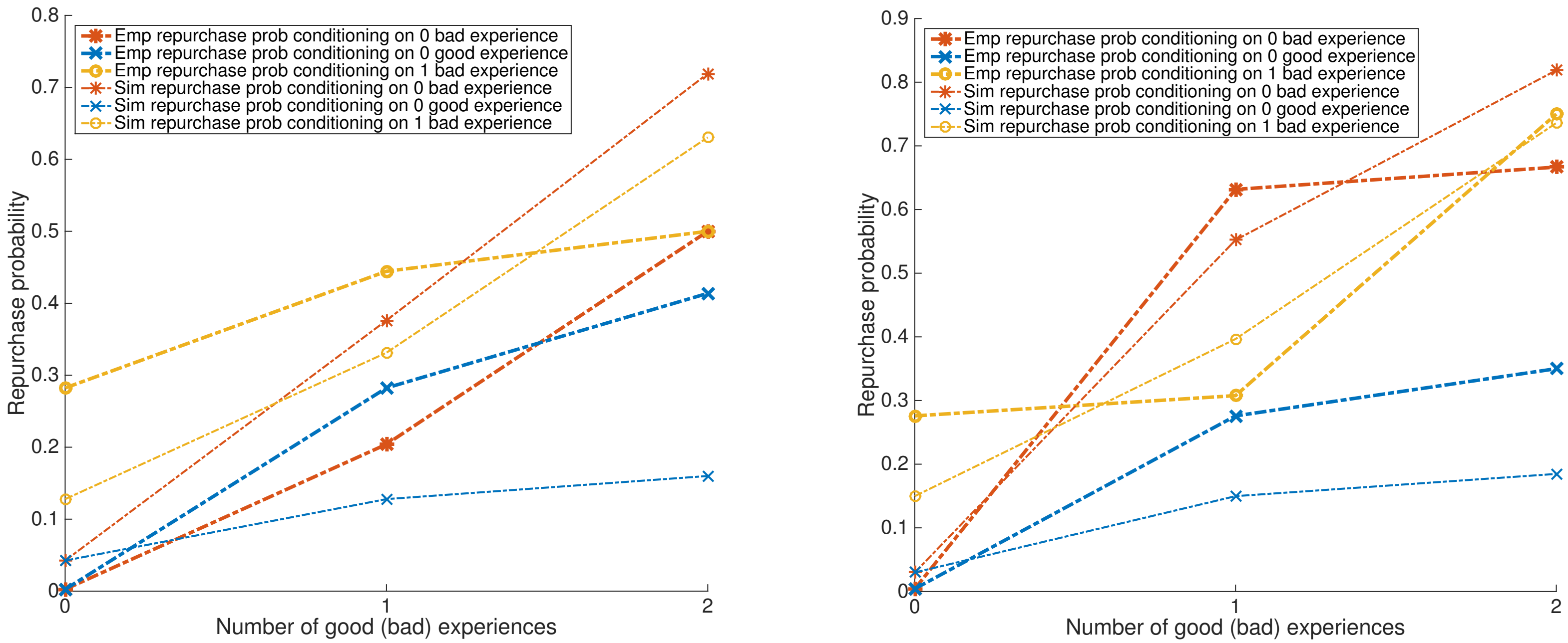


Note: This figure examines the goodness of fit of the simulated maximum likelihood estimates. The solid lines plot the empirical weekly market shares of the three product categories: the premium pile, the normal pile and the other sellers', which are computed using the households' purchasing records. The omitted category is for not purchasing any watermelon in a given week. The dotted lines plot the simulated market shares. Details on the simulation procedure are provided in Appendix E.4.

Appendix Figure 11: Model Fit: Simulated vs Actual Purchasing Patterns Conditional on Experiences

A. Sticker Markets

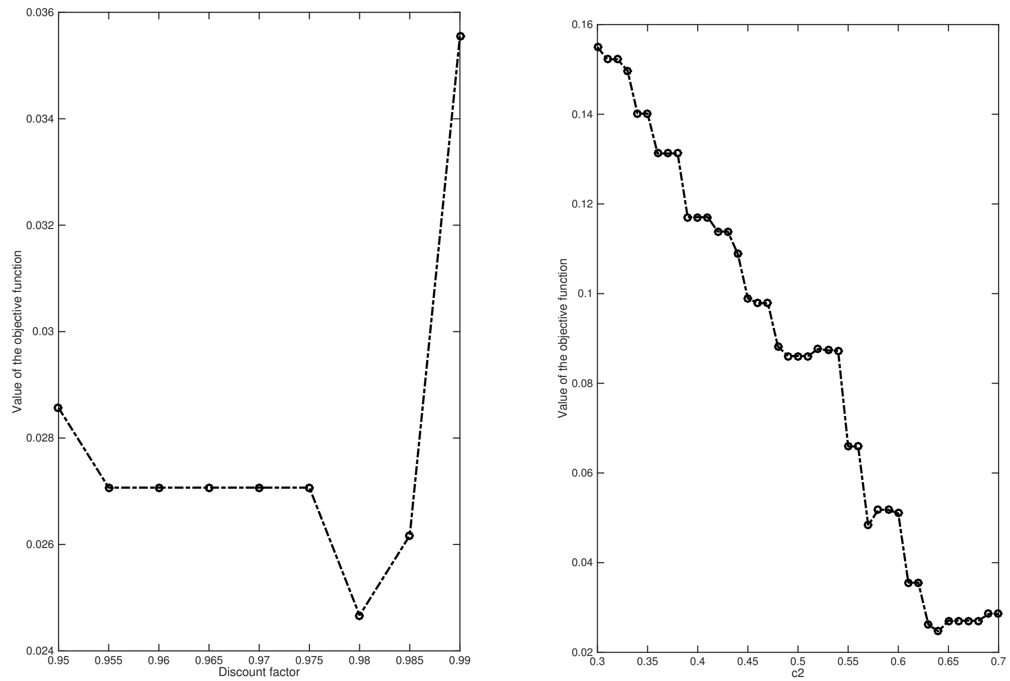
B. Laser Markets



Note: This figure examines the goodness of fit of the simulated maximum likelihood estimates. The solid lines plot the empirical repurchasing probabilities of the premium pile conditional experience combinations (see footnote of Appendix Table 11). The dotted lines plot the purchasing probabilities using the simulated data. Details on the simulation procedure are provided in Appendix E.4.

Appendix Figure 12: Value of the $v(\delta, c)$ for Various δ and c

$c = 0.64$ $\quad\quad$ $\delta = 0.98$



Note: This figure plots the value of the objective function for the minimum distance estimators as δ and c vary. The objective function is minimized using grid search and the optimal parameter values are $\delta = 0.98$ and $c = 0.64$.

.

Appendix Table 1: Balance Check for Baseline Community and Market Characteristics

	Label-less Non-incentive (1)	Label-less Incentive (2)	Sticker Non-incentive (3)	Sticker Incentive (4)	Laser Non-incentive (5)	Laser Incentive (6)	p-value (7)
Size measured in the number of housing units	1708.4	211.5	907.2	-301.6	445.2	-21.9	.781
	353.155	600.734	1047.796	458.423	797.797	731.985	.
Housing price (in RMB/meter^2)	8035.4	214.6	-715.9	919.6	451.7**	664.6	.092
	400.926	713.145	745.83	442.205	766.026	526.907	.
% of elderly	28.5	-5	8.5	2.5	-5	-4	.073
	4.537	5.431	6.021	6.094	5.38	5.845	.
Distance to the nearest supermarket (meter)	1320	620	380	195	10	275	.765
	369.248	525.674	517.161	504.439	431.946	496.356	.
Years since establishment	19.9	-5.7	3	-4.3	-2.6	-4	.708
	4.391	5.737	6.458	5.293	4.827	5.314	.
Number of competitors in the local market	3.9	-.3	.6	-.5	-1.3**	-.7	.18
	.407	1.363	.839	.709	.571	.654	.

Note: This table shows balance checks for main community and market characteristics across the treatment groups. Column 1 shows sample mean for the label-less non-incentive group (the omitted group). Columns 2 through 6 show the OLS regression coefficients of the other five group dummies. Column 7 shows the p-value for the Wald test of joint significance of the five coefficients. Standard errors are in parentheses.

Appendix Table 2: Balance Check for Baseline Seller Characteristics

	Label-less Non-incentive (1)	Label-less Incentive (2)	Sticker Non-incentive (3)	Sticker Incentive (4)	Laser Non-incentive (5)	Laser Incentive (6)	p-value (7)
Gender	.3	.3	.2	.2	.4*	.2	.591
	.153	.224	.226	.226	.216	.226	
Age	39.5	5.6	-1.3	3.9	1	.2	.604
	3.317	4.763	4.369	4.293	4.108	4.295	
Years of schooling	10.3	-.7	.2	.5	-.189	-.1	.921
	.871	1.456	1.1	.999	1.377	.999	
Number of years selling fruits	9.4	1.7	-.5	.5	-1.7	-2.3	.772
	1.759	3.21	2.194	2.694	2.617	2.416	
Number of years selling fruits at this location	7.4	3.7	-.4	1.4	-.9	-1	.73
	1.565	3.206	2.061	2.646	2.549	2.34	

Note: This table shows balance checks for sellers' characteristics across the treatment groups. Columns 1 shows sample mean for the label-less non-incentive group (the omitted group). Columns 2 through 6 show the OLS regression coefficients of the other five group dummies. Column 7 shows the p-value for the Wald test of joint significance of the five coefficients. Standard errors are in parentheses.

Appendix Table 3: Balance Check for Baseline Household Characteristics

	Label-less Non-incentive (1)	Label-less Incentive (2)	Sticker Non-incentive (3)	Sticker Incentive (4)	Laser Non-incentive (5)	Laser Incentive (6)	p-value (7)
Household size	3.4	.064	.624	.32*	.439	.683**	.132
	.186	.239	.318	.271	.315	.302	
% of elderly	.186	.017	-.051	-.042	-.057	.047	.352
	.075	.083	.089	.088	.078	.096	
% of female	.501	-.005	-.001	-.028	.012	-.004	.879
	.007	.013	.014	.029	.019	.013	
Household monthly income (in RMB)	5331.461	-464.525	-417.526	321.171	152.146	73.802	.67
	525.635	669.145	586.713	705.323	696.495	894.775	
Fruit consumptions as % of total food consumptions	31.133	5.95	.867	-1.171	-.182	-1.733	.187
	5.631	6.024	6.749	5.93	5.676	9.65	
Watermelon consumptions as % of total fruit consumptions	22.14	24.045**	11.36	23.329**	4.157	15.291	.045
	6.732	9.701	8.409	9.585	7.559	11.536	
Number of watermelons consumed per week	1.278	-.122	.079	.14	.094	-.005	.104
	.083	.12	.133	.091	.199	.114	.
Mostly buy watermelons from the local market (dummy)	.67	.186**	-.118	.08	.202***	.16**	.002
	.037	.085	.074	.113	.055	.078	
Mostly buy watermelons from nearby supermarkets (dummy)	.15	.018	.242	.14	-.014	.07	.096
	.121	.133	.143	.139	.13	.139	
Willingness to pay for quality (in RMB/Jin)	1.804	.095	.184	.186*	.135	.097	.388
	.053	.118	.112	.098	.102	.081	.

Note: This table shows balance checks for households' demographic characteristics across the treatment groups. Column 1 shows sample mean for the label-less non-incentive group (the omitted group). Columns 2 through 6 show the OLS regression coefficients of the other five group dummies. Column 7 shows the p-value for the Wald test of joint significance of the five coefficients. Standard errors are in parentheses.

Appendix Table 4: Quality Provision Using Household Satisfaction Ratings

	Ordered Probit: satisfaction rating from 1 to 5				Probit: dummy for the highest rating of 5			
	Non-incentive	Incentive	Laser	Sticker	Non-incentive	Incentive	Laser	Sticker
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
laser	0.534**	0.349			0.466*	0.335		
	(0.227)	(0.229)			(0.265)	(0.252)		
incentive			0.292	0.550**			0.431	0.860***
			(0.274)	(0.228)			(0.311)	(0.255)
Observations	127	125	93	83	127	125	93	83
Time Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓

Note: This table examines quality provision by treatment group using the data collected from the household. Quality is measured in terms of the satisfaction rating for watermelons purchased from the premium pile (Appendix C3 describes the details of the data). Each observation is a household-week purchase. The key explanatory variables are the group dummies. Panel A estimates an ordered probit model using the original self-reported satisfaction rating ranging from 1 to 5. Column B estimates a probit model for a dummy variable for the highest satisfaction rating of 5. Each column restricts to different subsamples in order to examine the interaction effect of the branding and incentive treatments (see footnote of Table 3. All regressions control for week fixed effects. Standard errors are clustered at the market level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 5: Pricing Behavior of Competitors

Dep var: daily market average price charged by competitors (in RMB/Jin)

	(1)	(2)
Incentive	0.002	0.001
	(0.015)	(0.015)
Sticker	-0.012	-0.012
	(0.009)	(0.010)
Sticker X Incentive	0.011	0.013
	(0.018)	(0.018)
Laser	0.007	0.007
	(0.014)	(0.014)
Laser X Incentive	0.014	0.016
	(0.022)	(0.022)
Constant	0.963***	0.963***
	(0.007)	(0.007)
Observations	2913	2913
Day Fixed Effects		✓

Note: This table examines the pricing behavior of the other sellers operating in these markets who are not included in the study sample across treatment groups. The dependent variable is the daily market average price charged by the other sellers (measured in RMB/Jin). The omitted group is the label-less non-incentive group. Column 2 controls in addition for day fixed effects. Standard errors are in parentheses, clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 6: Time Dynamics for Sales Quantity of the Premium Pile

	Laser		Sticker	
	(1)	(2)	(3)	(4)
Day	-0.576		-0.508	
	(0.385)		(0.803)	
Day X Incentive	1.598***		-0.309	
	(0.494)		(0.903)	
Week		-3.405		-3.589
		(2.635)		(5.670)
Week X Incentive		11.367***		-1.512
		(3.432)		(6.377)
Observations	971	971	976	976
Seller Fixed Effects	✓	✓	✓	✓

Note: This table shows the regression results of fitting a linear time model. The dependent variable is daily sales quantity of the premium pile, measured in Jin. Each observation is at the seller-day level. The key explanatory variable is the interaction term between the incentive treatment dummy and time (day or week). Columns 1 and 2 look within the laser groups; columns 3 and 4 look within the sticker groups. All regressions control for time and seller fixed effects. Standard errors are clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 7: Household Endline Perceptions

Dep var.: Willingness to pay for quality (in RMB/Jin)

	Un-branded		Sticker branded		Laser branded	
	(1)	(2)	(3)	(4)	(5)	(6)
Sticker	0.019	0.006	0.033	-0.001	0.138	0.080
	(0.031)	(0.034)	(0.067)	(0.072)	(0.098)	(0.103)
Laser	0.075**	0.062*	0.053	0.023	0.056	0.022
	(0.032)	(0.033)	(0.067)	(0.071)	(0.098)	(0.103)
Incentive	0.014	0.007	0.026	0.015	0.023	0.003
	(0.031)	(0.033)	(0.065)	(0.069)	(0.096)	(0.100)
Sticker X Incentive	0.027	0.037	0.108	0.136	0.055	0.099
	(0.044)	(0.045)	(0.093)	(0.097)	(0.136)	(0.139)
Laser X Incentive	0.020	0.039	0.034	0.067	0.311**	0.355**
	(0.044)	(0.046)	(0.094)	(0.097)	(0.138)	(0.141)
Observations	580	554	581	555	579	553
Household Baseline Controls		✓		✓		✓
Label-less non-incentive mean	1.115		1.218		1.489	
Std. dev.	(0.148)		(0.223)		(0.298)	

Note: This table examines the endline willingness to pay (WTP) for quality for households in different markets. To elicit the WTP, households were asked to consider a hypothetical situation where they see two piles of watermelons sold in the local market, one pile sold at 1 RMB/Jin and the other pile sold at a higher price with different branding technologies (un-branded, sticker and laser). Household reported the highest prices (in RMB/Jin) that they are willing to pay for the latter. This table regresses the self-reported WTP under different branding technologies on treatment dummies. The omitted group is the label-less non-incentive group. In addition, the even columns control for a set of household baseline characteristics, including household size, percentage of elderly, monthly income, average number of watermelons consumed per week reported in the baseline survey, and the baseline self-reported willingness to pay for quality (measured in RMB/Jin). Standard errors are clustered at the market level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 8: Spillover Effects to Sales of Peaches

Dep var: daily peach sales (in RMB)	(1)	(2)
Labelless incentive	-14.029	-23.361
	(29.712)	(28.593)
Sticker non-incentive	30.190	41.283
	(37.136)	(31.877)
Sticker incentive	-17.491	-23.835
	(29.814)	(28.966)
Laser non-incentive	-13.605	-30.616
	(28.715)	(29.367)
Laser incentive	42.153	29.056
	(36.721)	(34.705)
Constant	113.354***	9.557
	(26.722)	(61.867)
Observations	1312	1312
Seller (Market) Baseline Controls		✓

Note: This table shows the results of regressing daily sales of peaches in RMB, aggregated over all quality categories, on treatment group dummies. The omitted group is the label-less non-incentive group. Column 2 controls for the same set of seller and market baseline characteristics as that in Table 3. Standard errors are in parentheses, clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 9: Endline Reported Willingness to Pay for Different Branding Technologies

	WTP for sticker	WTP for laser
	(1)	(2)
Sticker	0.003	0.005
	(0.005)	(0.012)
Laser	-0.009*	0.042***
	(0.005)	(0.012)
Label-less	0.016***	0.104***
	(0.004)	(0.009)
Observations	59	59

Note: This table looks at sellers' endline self-reported willingness to pay for different branding technologies. The dependent variable is the self-reported willingness to pay for the two branding technologies, namely sticker and laser, and it is measured in RMB per watermelon. The omitted group is the label-less group. Standard errors are in parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 10: Price and Quality Dynamics

	Laser non-incentive group		Laser incentive group	
	Premium pile price Δ (over normal pile)	Premium pile quality (measured in sweetness)	Premium pile price Δ (over normal pile)	Premium pile quality (measured in sweetness)
	(1)	(2)	(3)	(4)
Day	0.000		0.000	
	(0.000)		(0.000)	
Check		0.042		0.062
		(0.043)		(0.040)
Constant	0.121***	10.181***	0.155***	10.551***
	(0.005)	(0.282)	(0.007)	(0.258)
Observations	483	119	488	117
Seller Fixed Effects	✓	✓	✓	✓

Note: This table examines the price premium and quality dynamics for sellers in the laser non-incentive and incentive groups. Price Δ is measured as the difference between the premium pile price and the normal pile price, in RMB/Jin. Quality is measured in sweetness. The key regressor of interest is a measure for time. All regressions control for seller fixed effects. Standard errors are in parentheses, clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 11: Purchasing Probabilities Conditional on Experiences

Num. of satisfactory exp (1)	Num. of non-satisfactory exp (2)	Total count (3)	Num. of purchases (4)	% of purchases (5)
Panel A. Households in laser markets				
0	0	1154	5	0.004
0	1	87	24	0.276
0	2	20	7	0.350
1	0	19	12	0.632
1	1	26	8	0.308
2	0	9	6	0.667
Panel B. Households in sticker markets				
0	0	1186	3	0.003
0	1	85	24	0.282
0	2	29	12	0.414
1	0	49	10	0.204
1	1	18	8	0.444
2	0	4	2	0.500

Note: This table summarizes the purchasing probabilities conditional on the number of satisfactory and non-satisfactory experiences. I stack together all household-week level observations that start with a given experience combination, and count the fraction among all those occasions in which a premium option was purchased by the household during that week. Column 3 counts the number of household-week observations that start with a given experience combination. Column 4 counts the number among all those occasions in which a premium watermelon was bought by the household during that week. Column 5 computes the fraction.

Interpretation: For households in the laser group, going from zero experience to one satisfactory experience increases the purchasing probability by about 63%, but having one additional satisfactory experience further increases the probability by only 3.5%. This pattern indicates a very noisy prior. However, the fraction of repurchasing also goes up with one bad experience. This is not surprising given that the compositions of households are different for the different cells. (This also points to the importance of allowing for richer observed and unobserved persistent heterogeneity among consumers.) Nonetheless, the difference in the repurchasing probabilities under (0,1) and (1,0) can be interpreted as the effect of learning because the total number of experiences is the same in these two cases, which controls for the composition effect. This difference is much more pronounced under laser than under sticker.

Appendix Table 12: Simulated ML Estimates for Market and Time Fixed Effects

Parameters	Baseline Model (1)		Direct Utility of Laser (2)		Correlated Learning (3)		Information Diffusion (4)	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Market FE 's	0.984	0.191	0.978	0.191	0.989	0.191	0.983	0.191
	0.538	0.153	0.535	0.153	0.536	0.153	0.533	0.153
	0.417	0.172	0.414	0.172	0.411	0.172	0.409	0.172
	-0.033	0.137	-0.037	0.137	-0.034	0.137	-0.037	0.137
	0.704	0.152	0.699	0.152	0.702	0.152	0.698	0.152
	-3.419	0.180	-3.426	0.180	-3.414	0.180	-3.420	0.180
	-0.610	0.139	-0.612	0.139	-0.611	0.139	-0.614	0.139
	-0.542	0.219	-0.537	0.219	-0.542	0.219	-0.547	0.219
	-0.203	0.321	-0.203	0.321	-0.200	0.321	-0.205	0.321
	0.342	0.199	0.340	0.199	0.344	0.199	0.338	0.199
	1.021	0.280	1.027	0.281	1.023	0.280	1.019	0.280
	-0.639	0.212	-0.639	0.212	-0.635	0.212	-0.641	0.212
	0.099	0.192	0.097	0.193	0.100	0.192	0.094	0.193
	0.197	0.267	0.199	0.267	0.198	0.268	0.194	0.268
	2.957	0.481	2.958	0.481	2.955	0.481	2.948	0.481
	-0.483	0.212	-0.478	0.212	-0.484	0.211	-0.487	0.212
	-1.479	0.272	-1.504	0.273	-1.486	0.272	-1.501	0.280
	0.051	0.229	0.033	0.230	0.055	0.229	0.050	0.230
	-1.122	0.226	-1.141	0.226	-1.116	0.226	-1.122	0.226
	-0.699	0.208	-0.715	0.208	-0.695	0.208	-0.700	0.208
	-0.274	0.200	-0.289	0.200	-0.271	0.200	-0.278	0.200
	-0.773	0.256	-0.797	0.257	-0.780	0.257	-0.790	0.257
	-0.803	0.223	-0.822	0.223	-0.798	0.223	-0.804	0.223
	-1.522	0.196	-1.545	0.196	-1.514	0.196	-1.521	0.196
	-0.764	0.255	-0.785	0.256	-0.773	0.254	-0.780	0.255
Mean	-0.242	0.223	-0.250	0.223	-0.242	0.223	-0.247	0.223
Time FE's	0.283	0.181	0.281	0.181	0.281	0.181	0.279	0.181
	0.314	0.182	0.311	0.182	0.309	0.182	0.304	0.182
	-0.185	0.167	-0.187	0.167	-0.191	0.167	-0.196	0.167
	-0.355	0.167	-0.357	0.167	-0.361	0.167	-0.367	0.168
	-0.409	0.182	-0.412	0.182	-0.415	0.182	-0.422	0.182
	-0.192	0.195	-0.196	0.195	-0.197	0.195	-0.205	0.196
Mean	-0.091	0.179	-0.093	0.179	-0.096	0.179	-0.101	0.179

Note: This table presents the simulated maximum likelihood estimates of the market and time fixed effects. Market fixed effects are estimated for all the 26 markets in the final analysis sample (see Appendix C.3 for details). Time fixed effects are estimated for week 2 to week 7 (relative to week 1, which is normalized to be 0).

Appendix Table 13: Optimal Policies under Different Parameter Values and Empirical Policies

δ	c	$\gamma_{\text{non-inc}}$ $\overline{\gamma}_{\text{non-inc}} = 0.4$	$m_{\text{non-inc}}$ $\overline{m}_{\text{non-inc}} = 0.142$	γ_{inc} $\overline{\gamma}_{\text{inc}} = 0.53$	m_{inc} $\overline{m}_{\text{inc}} = 0.178$	$v(\delta, c)$
0.97	0.64	0.41	0.25	0.46	0.28	0.027
0.98	0.64	0.41	0.25	0.48	0.28	0.025
0.99	0.64	0.46	0.28	0.48	0.28	0.036
0.98	0.63	0.44	0.25	0.48	0.28	0.026
0.98	0.64	0.41	0.25	0.48	0.28	0.025
0.98	0.65	0.41	0.25	0.46	0.28	0.027

Note: This table shows the optimal price premium and quality policies under different parameter values of δ and c. The optimal policies are found using grid search. The last column calculates the sum of the squared distance between the optimal policies and the empirical average policies.

Appendix Table 14: Welfare Analysis: Multiple Equilibria

	Oligopolistic competition under asymmetric information	
	High-quality equilibrium (1)	Low-quality equilibrium (2)
Quality and price		
Average quality of the pool ($\underline{\gamma}$)	0.300	0.300
Quality of the premium pile (γ_H)	0.440	0.400
Price premium of the premium pile (m_H)	0.170	0.120
No adjustment (disc. Σ of 5 seasons)		
Sales profits	28,292	24,920
Effort costs	7,863	5,225
Net profits (PS_{own})	20,429	19,695
Sales profits of other sellers	102,983	86,928
Effort costs of other sellers	31,973	21,210
Net profits of other sellers (PS_{other})	71,010	65,718
Expected maximum utility in RMB (CS)	484,279	457,123
Total surplus ($= \text{PS}_{\text{own}} + \text{PS}_{\text{other}} + \text{CS}$)	575,718	542,536
Ratio relative to baseline	2.188	2
With adjustment (disc. Σ of 5 seasons)		
Net profits (PS_{own})	14,695	14,602
Net profits of other sellers (PS_{other})	68,011	67,312
Expected maximum utility in RMB (CS)	361,737	352,805
Total surplus ($= \text{PS}_{\text{own}} + \text{PS}_{\text{other}} + \text{CS}$)	444,443	434,720

Note: This table compares the welfare consequences for the high-quality and low-quality symmetric equilibria for the case of oligopolistic competition under asymmetric information. Column 1 shows the high-quality equilibrium (as shown in Table 9). Details are explained under the footnote of Table 9.

Appendix Table 15: Heterogeneity Across Sellers in Pricing Behavior

Dep var: difference between the premium pile price and the market average price (in RMB/Jin)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gender (female=1)	-0.018								-0.007
	(0.011)								(0.013)
Age		0.001*							0.001
		(0.001)							(0.001)
Years of schooling			-0.002						-0.002
			(0.003)						(0.002)
Years of selling watermelons				0.002					0.001
				(0.001)					(0.001)
Housing price					0.008**				0.008*
					(0.003)				(0.004)
Pct of Elderly						-0.001*			-0.001
						(0.001)			(0.001)
Number of competitors							0.003		0.003
							(0.004)		(0.003)
Distance to supermarkets								-0.003	-0.002
								(0.006)	(0.006)
Constant	0.094***	0.044	0.108***	0.068***	0.024	0.125***	0.073***	0.091***	0.042
	(0.020)	(0.026)	(0.034)	(0.023)	(0.026)	(0.026)	(0.025)	(0.021)	(0.067)
Observations	1946	1946	1901	1946	1946	1946	1946	1946	1901
Group Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Day Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: This table examines heterogeneity across sellers in pricing behavior. Each observation is at the seller-day level. The dependent variable is the difference between the premium pile price and the market average price charged by other sellers in the same market. All regressions include group fixed effects, thus the difference represents heterogeneous pricing behavior conditional on the same treatment. All regressions control for day fixed effects. The explanatory variables are summarized in Panel B of Table 1. Standard errors are in parentheses, clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix Table 16: Heterogeneity across Sellers in Quality Provision

Dep var: premium pile sweetness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gender (female=1)	-0.307								-0.240**
	(0.205)								(0.108)
Age		-0.003							0.004
		(0.014)							(0.007)
Years of schooling			-0.082***						-0.059**
			(0.028)						(0.023)
Years of selling watermelons				0.014					-0.006
				(0.018)					(0.013)
Housing price					-0.032				0.005
					(0.023)				(0.038)
Pct of Elderly						-0.015*			-0.011
						(0.009)			(0.008)
Number of competitors							-0.077		-0.046*
							(0.055)		(0.026)
Distance to supermarkets								0.104	0.056
								(0.078)	(0.079)
Constant	9.122***	9.076***	9.906***	8.840***	9.200***	9.521***	9.313***	8.790***	10.175***
	(0.316)	(0.463)	(0.430)	(0.331)	(0.330)	(0.472)	(0.364)	(0.294)	(0.874)
Observations	238	238	226	238	238	238	238	238	226
Group Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Check Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: This table examines heterogeneity across sellers in quality provision. Each observation is at the seller-check level. The dependent variable is the sweetness of the watermelons in the premium pile. All regressions include group fixed effects, thus the difference represents heterogeneous pricing behavior conditional on the same treatment. All regressions control for check fixed effects. The explanatory variables are summarized in Panel B of Table 1. Standard errors are in parentheses, clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Appendix B. Theory Appendix

This appendix presents an alternative framework for analyzing sellers' quality provision under asymmetric information. The model is adapted from Klein and Leffer (1981). The primitives of the model follows the setup in Section 3.1 with one important distinction on the information structure and behavior of the consumers. The model assumes idiosyncratic experience shocks but complete information linkage across consumers. That is, consumers disseminate their experiences with one another after each period. With a continuum of consumers, the seller's choice of γ_H is perfectly known after each period. Therefore, the setup falls in the sphere of games with perfect public monitoring. To simplify notation, I normalize $P_N = P_W = 0$, and assume a unit mass of consumers ($M = 1$).

With this setup, a reputational equilibrium can be defined as follows:

Definition: A *reputational equilibrium* is a rational expectation Nash Equilibrium that consists of a pair (γ_H, P_H), where $\gamma_H > \underline{\gamma}$ and $P_H > 0$, such that:

1. Buyers with valuation $\theta > \hat{\theta}_1 = \frac{P_H}{\gamma_H - \underline{\gamma}}$ are willing to buy the good product at price P_H so long as all transactions up to date have yielded satisfaction rates bigger than or equal to γ_H. Once the satisfaction rate falls below γ_H, all buyers stop buying from the seller in future periods.

2. The seller is willing to provide quality γ_H for the good product rather than the minimum quality $\underline{\gamma}$ if his reputation is intact and chooses $\gamma_H = \underline{\gamma}$ otherwise.

A reputational equilibrium as defined above is a "trigger strategy" or "bootstrap" equilibrium in which a strictly positive mass of consumers are willing to buy the good product at the implicitly contracted quality-price premium, and the seller is wiling to provide $\gamma_H > \underline{\gamma}$ rather than deviating. For the first condition to hold, we need

$$\hat{\theta}_1 = \frac{P_H}{\gamma_H - \underline{\gamma}} < \overline{\theta} \tag{3}$$

so that there is positive demand for the good product in a reputational equilibrium. For the second condition, we need the seller's discounted future stream of profits to be greater than the current period gain of deviating. That is:

$$\frac{1}{\delta}(\overline{\theta} - \hat{\theta}_1)(P_H - C(\gamma_H)) \geq (\overline{\theta} - \hat{\theta}_1)P_H$$

Re-arranged, we get:

$$P_H \geq \frac{C(\gamma_H)}{\delta} \equiv P_{min} \tag{4}$$

where P_{min} is called the minimum quality-assuring price, which arises because of asymmetric information.

Combining Equation (3) and (4), we can derive the following proposition:

Proposition B.1: (*On the existence of reputational equilibrium*) If $\frac{\overline{\theta}(\gamma_H - \underline{\gamma})}{C(\gamma_H)} < \frac{1}{\delta}$, a reputational equilibrium doesn't exist. If $\frac{\overline{\theta}(\gamma_H - \underline{\gamma})}{C(\gamma_H)} \geq \frac{1}{\delta}$, any $\gamma \in (\underline{\gamma}, 1]$ can be supported as a reputational equilibrium.

Note that there is always a pooling equilibrium in which no buyer buys anticipating that the seller exerts minimal effort, and the seller exerts minimal effort anticipating that no buyer will wish to buy in the future.

In light of Proposition 1, there are two broad explanations for the absence of quality premium at baseline. First, a reputational equilibrium doesn't exist. If the cost of supplying higher quality is very high relative to consumer's willingness to pay for quality, and because of imperfect information the price must be even higher than the cost in order to incentivize the seller to exert effort. How high the minimum quality-assuring price must be depends on seller's discount rate and the demand conditions, and it could well be that the minimum quality-assuring price exceeds the increase in consumer surplus of purchasing the higher quality product. If so, products of higher than the minimum quality will not be demanded and won't be supplied. In such a case, a reputational equilibrium doesn't exist.

A second explanation for the lack of quality differentiation is that there exist multiple equilibria, but the markets at baseline are stuck in a pooling one. In reality, we may expect that the reputational equilibrium that maximizes a seller's profits is more likely to prevail instead of the pooling one. Thus a pure coordination failure may not be a very appealing explanation. With uncertainties or misinformation about costs, pooling could be more likely. I discuss this next.

Uncertainties and the Role of Advertisement

Klein and Leffler (1981) consider the role of advertisement in the form of initial sunk cost investment in a world where consumers are uncertain about the cost conditions. The key idea is the logic of forward induction that refines the set of equilibria.

If consumers are uncertain about the underlying cost of producing high quality, then a high price might indicate either a high quasi-rent or a high cost. To illustrate this idea, suppose that the cost is linear in quality, $C(\gamma_H) = c(\gamma_H - \underline{\gamma})$. Suppose consumers perceive the quality gradient of the cost function to be $\tilde{c}$, which is higher than the true cost c. Then the minimum quality assuring price is $\frac{\tilde{c}}{\delta}(\gamma_H - \underline{\gamma})$. If $\frac{\overline{\theta}}{\tilde{c}} < \frac{1}{\delta} < \frac{\overline{\theta}}{c}$, then consumers are unlikely to be swayed by the high price.

In this context, upon seeing a conspicuous initial sunk cost investment of W and observing a price p_H, assuming that sellers play the profit maximizing reputational equilibrium, the forward induction argument implies that

$$\frac{1}{1-\delta}(\overline{\theta} - \frac{P_H}{\gamma_H - \underline{\gamma}})(P_H - \tilde{c}\gamma_H) - W \geq 0 \Rightarrow \tilde{c}\gamma_H \leq P_H - \frac{W(1-\delta)(\gamma_H - \underline{\gamma})}{(\gamma_H - \underline{\gamma})\overline{\theta} - P_H} \tag{5}$$

where $\tilde{c}$ is the perceived cost. This condition says that consumers think the discounted sum of future quasi-rents must be large enough to justify the initial sunk cost investment because otherwise, sellers would not have made that investment. Hence, this argument puts an upper bound on $\tilde{c}$. As W increases, $\tilde{c}$ can be made arbitrarily small.

Supposing that the forward induction argument works, then sellers with real cost $c < \delta\overline{\theta}$ could invest in W to shift $\tilde{c}$ right to the cutoff to ensure that the equilibrium played is the reputational equilibrium and earn positive profits (after taking into account the sunk cost investment).

A natural question arises: why have sellers not already invested in such an advertisement technology to ensure that the equilibrium played is the reputational equilibrium instead of the pooling one? First, such an advertisement device may not be available and consumers are not necessarily swayed by just any form of money-burning activity. More importantly, the model does not speak about equilibrium switching dynamics. In reality, it is hard for sellers and consumers to simultaneously coordinate on a new equilibrium upon observing a sunk cost investment.[40] The model in Section 3 incorporates consumer learning dynamics, though consumers are modeled as being naive. The two frameworks offer two explanations for the lack of quality differentiation at baseline: for the former, the problem is a coordination failure, whereas for in the latter, the problem is pessimistic prior beliefs, which become self-confirming.

[40] This model has households guessing about the seller's costs and draw an inference. To get learning going, however, the households need to mis-estimate the true cost structure and some learning rule needs to be specified.

Appendix C. Sampling, Recruitment and Data Issues

C.1 Sampling and Recruitment

Screening markets. A baseline census of all gated communities and local markets in the city of Shijiazhuang was conducted in April 2014. Basic demographic and socioeconomic information was collected for all gated communities and local markets in the city. The information was used for stratifying the randomization (see below). To minimize travel costs, I restrict the study area to three of the five districts in the city. To ensure a more homogeneous study sample, I restrict the sample to local markets that are present all year long (excluding those temporary markets, called "Ji" in Chinese, which only form on special dates in a month) and that house more than one fruit seller. In total, 130 local markets fit these screening criteria.

Recruiting sellers. One seller in each local market was selected for an expression of interest survey, and one seller in each market was selected following this sequential procedure:

(1) The seller must be selling multiple fruits in the summer, including watermelons. This is also to minimize baseline heterogeneity for power concerns. Pure watermelon sellers sometimes harvest watermelons directly from the fields, whereas most multi-fruit sellers source watermelons from the wholesale market. There are two big wholesale markets in the city. Most sellers in the study sample source from the same wholesale market.

(2) Among all sellers that meet the first criterion, the one located closest to the entrance was selected in order to facilitate the logistics of the labeling service. If there are multiple entrances or multiple sellers at the same entrance, the one with the largest store was selected. For this reason, the sample sellers tend to be larger than the other sellers in these markets. However, the main empirical analysis focuses on comparisons across sellers in different markets (since the randomization is at the market level) rather than across sellers in the same market.

Surveyors approached the selected sellers as agents from a marketing research company and conducted an expression of interest survey. In particular, sellers were asked whether they would be interested in participating in a two-month field research project that studies people's fruit consumption patterns in the summer. They would be paid 100 RMB per week for participating in the study, and in return, they would need to agree to follow these procedures: (1) record daily sales information for watermelons and peaches; (2) experiment with differentiating watermelons by quality at sale for some (unspecified) period of time. Sellers were told that they were free to set prices and would not be interfered in any other aspect of their business. The goal of conducting the expression of interest survey prior to randomization is to minimize attrition, and in particular, differential attrition across different treatment arms. For that purpose, sellers were also told that they might get a free labeling service for the higher quality watermelons, either in the form of laser or sticker. Sellers were ranked by their interest to participate in the study and the top 60 were selected to be in the final study sample.

Recruiting households. For five (out of ten) randomly chosen markets in each treatment group, a random sample of households from a nearby gated community were recruited. Some markets span multiple communities. In those cases, the community was chosen based on the following sequential elimination procedure:

(1) Exclude very small communities with fewer than 5 buildings (dan-yuan).

(2) Among the remaining ones, restrict to those located closest to the sample seller.

(3) Select the largest community (measured in terms of housing units) among those that satisfy the above two criteria.

During the recruitment, surveyors put up a table at the gate of the community and approached residents as representatives from a marketing research company. To ensure that the household sample represented a good mix of the population, surveyors went in during the late afternoons when local population flow through the gate is the largest. The target was to recruit an equal mix of people aged above 60 years old, between 40 and 60, and below 40. The actual age distribution is close to the target. Each participating household needed to record all of the family's fruit purchasing and consumption experiences over a two-month period, and in return, the household would receive a fruit coupon of 10 RMB at the end of every week. The coupon could be redeemed at the sample seller's store for any fruit purchases. The original target was to recruit 25 households in each of the 30 selected communities. Unfortunately, for three communities, the gatekeepers obstructed the recruiting process, and as a result, they were dropped from the household sample. The three communities look similar to the others on observable baseline characteristics. In total, 675 households in 27 communities were successfully recruited for the study.

C2. Daily Sales Recording: Issues and Cleaning Procedure

The 60 sample sellers were asked to record down their daily sales information for watermelons and peaches on a daily sales recording sheet. For each transaction, sellers were asked to record the fruit type (watermelon or peach), sales quantity (in Jin), sales values (in RMB), and the corresponding quality category, premium or normal, if the sold fruit is watermelon. Sellers were also asked to distinguish between different breeds of watermelons. For all the empirical analyses, I focus on the most popular breed, called "Jingxin" in Chinese. Sales of all the other breeds constituted less than 2% of the total recorded sales.

Omissions and errors in recording were unavoidable, and occasionally sellers had to lump several sales together if they happened around the same time. It would be of concern if for some reason the noise in recording differs systematically across the treatment groups. To check this possibility, a second source of sales information was collected starting from mid-August. In particular, besides the transaction-level records on each day, sellers were also asked to recall the total sales quantity of the previous day. As a first pass, the difference between the self-recalled and the recorded total sales quantity does not differ significantly across the treatment groups.

A related concern is that there might be differential recording noises by quality categories across the treatment groups even though the aggregate sales of the two piles do not differ. To examine this concern, I compare the daily sales quantity of the premium pile recorded by sellers with that inferred from the surveyors' records. In particular, on each day before surveyors carried out the branding service, they counted the number of branded watermelons left from the previous day and the number of newly branded ones. Using this information, I could back out the number of branded watermelons sold on a given day. While the timing difference between the branding service and the collection of the recording sheet introduces some additional noise, the finding that the correlation between the two measures does not appear to differ between the laser and sticker groups serves as a first pass and alleviates some of the concerns for differential recording noises across groups that may drive the empirical results.

C3. Household Panel Purchasing: Issues and Cleaning Procedure

Household recording sheets were distributed and collected weekly. Since the gated communities spread throughout the whole city, it was not possible to get all households to turn in their recording sheets at one central location. Therefore, to ease the logistical work, one household in each gated community was designated as the household in charge, and they took responsibility for collecting and distributing the recording sheets for the rest of the participating households in the community. Surveyors then collected the forms from the household in charge. While this procedure greatly reduced the logistical obstacles, it also made it difficult to spot and correct recording errors on time.

Broadly speaking, there are two major issues with the household data: first, for some transaction records, one or more of the following information could be missing: the date of purchase, fruit type, purchase place, purchase quantity, purchase value, whether the fruit purchased has any labeling or not, and the self-reported satisfaction rating; second, the records are missing for some households in some weeks, either due to family travels or other reasons that led to failure of collection. The latter is less of an issue than the former since weekly fruit coupons were distributed upon collecting the recording sheets, which gave households an incentive to turn in their forms on time.

The following table lists the percentage of watermelon transaction records with missing information of purchase place, labeling dummy and satisfaction rating in each of the treatment groups.

Group	Number of records	% of missing place	% of missing labeling	% of missing satisfaction
Label-less non-incentive	574	.056	.230	.129
Label-less incentive	710	.107	.211	.074
Sticker non-incentive	868	.021	.131	.101
Sticker incentive	894	.102	.199	.044
Laser non-incentive	820	.194	.224	.241
Laser incentive	650	.073	.190	.126

Missing information poses a serious problem in examining the dynamic purchasing and learning patterns because missing information would be treated as non-purchase. For example, in computing the number of watermelons purchased from the sample sellers in a given week, if the household failed to fill in the purchase place information, those purchases would not be counted even if they were actually made from the sample seller. The problem arises similarly for counting the number of premium pile purchases if the labeling information was missing.

I follow the below procedure to clean the household data and infer some of the important information missing in the records for watermelon purchases:

1. For missing purchase place, I code it up using the mode of purchase place for the household-fruit type. For example, if we observe in the data that the household mostly buys watermelons from the sample seller (i.e. the seller in our study sample), then I coded watermelon purchases with missing purchase place information as made from the sample seller.

2. Merge the household data with sellers' and surveyors' daily records and use the price information to infer missing pile (i.e. labeling) information for watermelons bought at the sample seller's store. For example, if a household recorded one watermelon bought from the sample seller's store on July 19th at a price of 1.2 RMB/Jin, which is the unit price charged for the premium pile watermelon on that day, I code the purchase as being made from the premium pile. In cases where the date information is missing, I compared the recorded price with the average weekly price charged by the sample seller for each pile. If the difference between the recorded price and the average weekly price for a given pile is smaller than 0.05 RMB, then I consider the purchase as being made from that pile.

3. I drop households that submitted fewer than 5 weekly records. Quite a number of the households submitted fewer than 8 weekly records. However, a couple of weeks of missing recording sheet could be due to travel, in which case it is conceptually correct to treat the number of watermelons consumed for that week as 0. 102 households were dropped. In particular, more than 15 households (out of 25) in one community were dropped as a result of turning in fewer than 5 weekly records. This could be due to the incompetency of the household in charge. I further dropped that community from the analysis. The empirical results are robust to this sample restriction.

The final sample consists of 4,309 watermelon purchase records from 573 households in 26 communities. The baseline characteristics of these 573 households are summarized in the table below. In general, they look very similar to the full sample (see Table 1).

The final analysis sample contains no missing place information and no missing pile information for purchases made at the sample seller. Satisfaction rating is missing for 11% of

	Observations	Median	Mean	Std. Dev
Household size	572	4	3.783	1.377
% of elderly	572	0	.174	.275
% of female	572	.5	.502	.154
Household monthly income (in RMB)	568	4000	5117.077	3143.528
Fruit as % of total food consumption	525	30	31.714	17.744
Watermelon as % of total fruit consumption	545	30	35.744	25.213
Num. of watermelons consumed per week	572	1	1.26	.653
Mostly buy watermelons in local market	573	1	.771	.42
Mostly buy watermelons in supermarket	573	0	.239	.427
Willingness to pay for quality (RMB/Jin)	551	2	1.94	.312

the purchases. Overall, 30.7% of the recorded purchases are made from the sample sellers' stores, and 57.2% are made from other sellers located in the same local market and nearby supermarkets; the rest are from other places. On average, each household buys 1.1 watermelons per week and the median is 1. These descriptive patterns all look similar to those for the full sample.

C.4 Household endline survey

Household endline survey was distributed and collected together with the last week's recording sheet. Overall, 10% of the households did not turn in the last week's recording sheet. Characteristics of households with missing endline data look similar to those who turned in and do not differ across groups. To examine changes in perceptions, the same question to elicit willingness to pay for quality was asked again, but this time for watermelons under three different branding technologies. Specifically, households were asked to compare two piles of watermelons, one of ordinary quality at 1 RMB/Jin and the other of premium quality with laser branding, sticker branding and no branding (label-less) respectively. For each scenario, households were asked to indicate the highest price they were willing to pay for the premium option. The reference price for the normal pile was also different from the baseline in order to match the actual average market price at the time when the endline survey was conducted.

Appendix D. Additional Tests, Data Summary and Regression Analysis

D.1 Sorting Ability Test

To formally establish information asymmetry in this setting, I conducted a sorting ability test with 30 fruit sellers in 30 different local markets in the city. Each of them was asked to sort 10 watermelons into two piles: one for high quality and one for low quality. Specifying a fixed number of watermelons for each pile may wash out differences between skilled and unskilled subjects, while not doing so can lead to trivial sorting. In practice, enumerators did not specify a fixed number of watermelons in each pile but suggested a range instead: the maximum and minimum for each pile are set to be 7 and 3 respectively. On average, sellers sorted 4.4 watermelons to the premium pile and consumers sorted 3.5. The watermelons were randomly picked by surveyors from the sellers' stores with no obvious distinguishable differences in outlook. The same test was repeated with 5 randomly chosen local consumers in each market. Finally, quality was measured using a sweetness meter. A baseline blind tasting test shows that sweetness strongly correlates with consumer's taste: among 210 consumers who were asked to compare two watermelons of high and low sweetness measures, 97% preferred the sweeter one.

D.2 Sales Dynamics Between Incentive and Non-incentive Groups

Panel B of Table 3 shows that the incentive led sellers to provide higher quality for both sicker and laser groups. Given this, we expect that over time as consumers experience the product, the incentive groups should outperform their non-incentive counterparts, especially under laser label where learning is salient (see Table 2). To investigate the time dynamics, I fit a linear time model that plots the group average quantity over time in the raw data):

$$\text{Premium Quantity}_{it} = \alpha + \beta_1 \text{Time}_t \times \text{Incentive}_i + \beta_2 \text{Time}_t + \gamma_i + \epsilon_{it} \tag{6}$$

where quantity on the LHS is at the seller-day level. Time is either day or week. I run this separately for sticker and laser groups, controlling for seller fixed effects. Appendix Table 6 shows the results. The significant positive coefficients for the interaction terms between the incentive treatment dummy and time for the laser group suggest that as consumers learn about the underlying quality over time, higher efforts could pay off. On the contrary, we do not see such a time pattern for the sticker group, which is consistent with the previous finding that consumers' beliefs update more slowly under the old technology.

D.3 Changes in Household Endline Perception

To examine whether sellers in the incentive group are endowed with higher beliefs than their non-incentive counterparts at the time when the incentive was lifted, ideally we would like to have data on market perception at each point in time. While this information is not directly available, households' perceptions elicited in the endline survey can be suggestive (Appendix C.4 explains how this was conducted). Appendix Table 7 regresses households' self-reported willingness to pay under different branding technologies on treatment group dummies. The omitted group is the label-less non-incentive group and the even columns control in addition for household baseline characteristics. We see that the willingness to pay is the highest under laser, and more importantly, it is the highest for households in the laser incentive markets. Households in the sticker incentive markets also appear to be willing to pay more under sticker, but the estimate is noisy and the magnitude is much smaller.

Appendix E. Structural Appendix

E.1 Simulated ML estimation and standard error calculation

The learning model is estimated using the simulated maximum likelihood method. The likelihood of household i for making an observed sequence of purchases can be computed as:

$$l_{ni} = \prod_{t=1}^{T}\prod_{j=0}^{3}\mathbb{E}\left(\mathbb{1}\{V_{imjt}+\epsilon_{imjt} > V_{imkt}+\epsilon_{imkt}, \forall k \neq j\}\right)^{d_{imjt}} = \prod_{t=1}^{T}\prod_{j=0}^{3}\left(\frac{\exp(V_{imjt})}{\sum_{k=0}^{3}\exp(V_{imkt})}\right)^{d_{imjt}}$$

The random effects η and ξ are assumed to follow independent distributions of $\log(\mathcal{N}(m(\eta),\sigma(\eta)))$ and $\mathcal{N}(m(\xi),\sigma(\xi))$, the average likelihood function for each household, $\tilde{l}_i$, can be computed by averaging l_{ni} over a large number of draws. The final objective function is obtained by multiplying $\tilde{l}_i$ across all households and taking log. Standard errors are computed using the outer product of gradient estimate of the asymptotic covariance matrix.

300 2-dimensional Halton draws are used to simulate the random effects η and ξ. This is drawn once and fixed throughout the optimization routine. The length for each draw of η is $\text{Nhh}_l + \text{Nhh}_s$ and that for ξ is $\text{Nhh}_l + \text{Nhh}_s + \text{Nhh}_{ll}$, where $\text{Nhh}_l, \text{Nhh}_s$ and Nhh_{ll} denote the number of households in the laser, sticker and label-less markets.

If I estimate the model unconstrained, a_0 turns out to be slightly negative. This is because of the small initial market share of the premium option observed in the data. In other words, the data suggests a very pessimistic prior as viewed through the lens of this model. In principle, including a product-specific constant (in this case, the random effects η and ξ) in the utility specification could help to alleviate the constraint. However, it doesn't in this case–the estimate of a_0 hits the zero boundary regardless of whether I include the random effects or not. One reason is that the data also points towards a very small a_0 in order to match the fast switching response: consumers with 1 or 0 good experience display very different repurchasing behavior; the largest possible difference the model could allow is when $a_0 = 0$, together with a small b_0. It's clear from this discussion that identification of the prior parameters relies on the dynamic purchasing patterns, which can be demanding given the short panel we have. An alternative approach is to constrain the prior mean to be the same as the existing option (q) and estimate the sum of $a_0 + b_0$, the smaller the value the larger the variance. The results (not shown) similarly indicate a much more stubborn prior under sticker than under laser. However, a simple likelihood ratio test strongly rejects the alternative model at 1% level (against the baseline model in Column 1 of Table 7).

In actual estimation, constraints are set for $a_0(\geq 0), b_0(\geq 0)$ and $q(0 \leq q \leq 1)$. Since the estimated a_0's hit the boundary, the asymptotic theory does not give a valid approximation for the standard errors in such a case. I set a_0's to be 0 and calculate the standard errors for the remaining parameters. In particular, I compute the outer product of gradients (OPG) estimate for the asymptotic covariance matrix:

$$\hat{V} = \left(\frac{1}{\text{Nhh}} \sum_{i=1}^{\text{Nhh}} \nabla_\theta ln(\tilde{l}_i(\hat{\theta})) \nabla_\theta ln(\tilde{l}_i(\hat{\theta}))^T \right)^{-1}$$

where Nhh is the total number of households in the sample (across all groups). $\tilde{l}_i$ is the average likelihood function for each household i for observing a sequence of purchasing decisions under parameters $\hat{\theta}$ (averaging across the draws for η and ξ). The gradient is computed numerically by taking small ϵ (0.00001) perturbations for each element of $\hat{\theta}$ and computing the difference in the log average likelihood function.

Under regularity conditions, the OPG estimator provides a consistent estimate of the asymptotic covariance matrix. Standard errors are given by the square root of the diagonal element of $\frac{1}{\text{Nhh}}\hat{V}$.

E.2 Alternative prior specifications

The demand model assumes a Beta prior distribution. An alternative approach would be to specify a Dirichlet's prior, the multivariate generalization of Beta distribution. However, doing so rules out updating among close-by categories. In the actual recording sheet, ratings from 1 to 5 stand for very bad, bad, ordinary, good, and very good. An alternative is to classify {4,5} to be satisfactory. However, the empirical satisfaction rate is as high as 85% for the undifferentiated pile for the alternative definition, and there is no distinguishable difference across the treatment groups. On the other hand, classifying 5 to be satisfactory results in a 30% satisfaction rate for the undifferentiated pile, and the rate is significantly higher for the incentive groups than for the non-incentive groups, consistent with the objective sweetness measure. These patterns suggest that consumers may be more discerning on the "very good" rating, thus the data speak for classifying 5 to be satisfactory as opposed to both 4 and 5. Finally, it is also worth mentioning that the self-reported satisfaction rating could well be subjective (i.e., household-specific). Classifying good and bad experiences as being above and below (or equal to) the median of each household produces qualitative similar results.

E.3 Supply side calibration

To construct the hypothetical *average market*, I pool together the194 households in the laser sample and scale up the market size by 4.5 to match the initial period's total sales quantity.

The empirical average quality is calibrated using the empirical satisfaction rate for watermelons purchased from the premium pile, which is 0.40 for laser non-incentive group and 0.53 for the incentive group. $\underline{\gamma}$ is calibrated using the satisfaction rate for watermelons purchased from the non-treated sellers, which is 0.3. $\overline{m}_H$ is the price difference (in RMB/Jin) between the premium pile and normal pile averaged across all sellers in a given group over time. The average price premium is 0.178 for the incentive group and 0.142 for the non-incentive group.

One concern of looking just at the average behavior is that the average could mask significant individual heterogeneity across sellers. Appendix Table 15 and 16 examine sellers' price and quality choices. Generally speaking, while policies do vary across sellers, most seem to be on dimensions related to the demand conditions, which are already captured in the current framework. Nonetheless, there could be other important dimensions of individual heterogeneity that are not observed in the data but that affect a seller's incentive to provide quality. The empirical model focuses on the effects of the demand conditions and therefore abstracts away from other aspects of individual heterogeneity.

E.4 Simulating dynamic demands and beliefs evolution

With the demand system estimates, for any given set of supply-side policies (i.e. prices and quality measured in terms of satisfaction rate), I follow the procedure below to forward simulate the demands for each product in each period as well as each period's experience realization, which is used to compute the posterior beliefs:

1. Draw 300 matrices of size Nhh $\times$ 2, where Nhh is the number of households in the sample from a two-dimensional Halton Sequence. Denote as H_s, for $s = 1 \dots 300$. Inverse cumulative normal is taken for each element of the matrices. The first column of each H_s is used for generating the random effect η, which is $\eta_s = \exp(\hat{m}(\eta) + \hat{\sigma}(\eta) H_s(:, 1))$; and the second column of each H_s is used for generating the random effect ξ, which is $\xi_s = \hat{m}(\xi) + \hat{\sigma}(\xi) H_s(:, 2))$. Thus, this creates 300 draws of η and ξ for each individual in the sample, and these draws are fixed throughout the simulation exercises.

2. Replicate the characteristics and the generated random effects of the Nhh households

1000 times. Characteristics include all individual baseline characteristics that are included in the mean utility function as well as the fixed effect estimate and empirical prices for the market an individual is associated with. For most simulation exercises (except for counterfactual exercises that require re-optimizing over price), I use the empirical average weekly prices for each market as seen in the data. The price vector faced by each household (and the 1000 replicas) in each period includes the price for the premium pile, the price for the normal pile and the average price charged by the other sellers' in the market. Let RepN=1000.

3. Draw two random uniform matrices, M_1 and M_2, of size Nhh $\times$ RepN by T $\times$ Sea. T indicates the number of weeks in one season, $T = 7$. Sea is the number of seasons (or years) that are simulated forward to. M_1 is used for determining purchasing decisions and M_2 is used for generating experience realizations. These matrices are drawn once and fixed for each iteration on s.

4. For each s, forward simulate the purchasing decisions and beliefs:

    (a) Begin with time t, compute $V_{imjt}, \forall i, j$ using individual i's time t posterior μ_{im1t}. Compute the probability of purchasing product j $\pi_{imjt} = \frac{\exp(V_{imjt})}{\sum_{k=1}^{J} \exp(V_{imkt})}$.

    (b) Generate purchasing decision d_{imjt} using π_{imjt} and the tth column of M_1.

    (c) Conditioning on purchasing $d_{im1t} = 1$, generate experience realization (dummy) using the satisfaction rate (given) and the tth column of M_2.

    (d) Update beliefs to $\mu_{im1,t+1}$ using the formula for Beta posterior.

    (e) Repeat (a) to (d) for $t = 1, \ldots,$T $\times$ Sea.

    (f) Sales quantity for each product option, market shares, and average beliefs can be computed for each period. Divide by RepN when necessary to get the expected number for the original Nhh households.

5. Repeat Step 4 for all $s = 1 \ldots 300$, and compute the averages for the measures in (f).

For supply-side analysis and welfare calculations, I follow the same procedure described above to first forward simulate the demand and beliefs evolution. After that, sellers' profits (of sample and non-sample sellers) and consumer surpluses can then be computed along the simulated paths.

E.5 Welfare Calculations

Without information problems, total consumer surplus (in RMB) can be computed using the standard log sum formula, which is the total discounted sum of expected maximum utility scaled by the price coefficients:

$$E(CS) = \frac{1}{T}\sum_{t=1}^{T}\delta^{t-1}\sum_{i=1}^{Nhh}\left[\frac{1}{\alpha_0 + \alpha_1 \text{Highinc}_i}\log\left(\sum_{j=1}^{J}\exp(V_{imjt}(\eta_i, \xi_i))\right)\right]$$

And producer surplus is the discounted sum of expected net profits:

$$E(PS) = \sum_{t=1}^{T}\delta^{t-1}\sum_{k\in\mathcal{K}}\sum_{i}^{Nhh}\frac{\exp(V_{imkt}(\eta_i, \xi_i))}{\sum_{j=1}^{J}(V_{imjt}(\eta_i, \xi_i))}\times(P_{mkt} - P_{\text{wt}} - C(\gamma_{kt}))$$

where $\mathcal{K}$ is the set of product options that a seller offers, either in just a normal pile (in which case $\gamma_k = \underline{\gamma} = 0.3$ and $C = 0$) or in both a normal pile and a premium pile, depending on the counterfactual exercises. Results are averaged over a large number of draws for the random effects η and ξ.

With information problems, consumer surplus takes a more complicated form because beliefs under which purchasing decisions are made are different from the truth. Leggett (2002) develops a solution to this problem for Type-I extreme value random utility errors with constant marginal utility of wealth. In particular, for consumer i in a given period t, the expected maximum utility (in RMB) is given by:

$$E(CS_{it}) = \frac{1}{\alpha_0 + \alpha_1 \text{Highinc}_i}\left[\log\left(\sum_{j=1}^{J}\exp(V_{ijt}(\mu_{ijt}))\right) + \sum_{j=1}^{J}\tilde{\pi}_j(V_{ijt}(\gamma_j) - V_{ijt}(\mu_{ijt}))\right]$$

where

$$\tilde{\pi}_j = \frac{\exp(V_{ijt}(\mu_{ijt}))}{\sum_{k=1}^{J}\exp(V_{ijt}(\mu_{ijt}))}$$

The second term in the outer bracket takes into account the fact that purchasing decisions are made under the current beliefs μ_{ijt}, whereas the true underlying quality is γ_j. To calculate welfare under asymmetric information, I forward simulate market evolution for given quality and price (see E.4) and use the adjusted log sum formula to compute the consumer surplus along the adjustment path.