Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation*

Jie Bai (Harvard Kennedy School) †

September 2024

Abstract

Quality provision is often low in many developing markets, and firms commonly lack a reputation for quality. This paper examines this issue both theoretically and empirically in the context of retail watermelon markets in China. I first demonstrate the existence of significant asymmetric information on quality between sellers and buyers, as well as the absence of a quality premium at baseline. To explain this, I develop a theoretical model that highlights the role of consumer beliefs and costly signaling in influencing sellers' reputation incentives. I then conduct an experiment by randomly introducing two signaling technologies into different markets: a cheap sticker label and a more expensive laser-cut label. Consistent with the theoretical predictions, the laser label induces sellers to offer higher quality, resulting in increased sales and profits, while the sticker label fails to achieve the same effect. Using the experimental variation, I estimate an empirical model of consumer learning to uncover the underlying evolution of beliefs. The results show that pessimistic beliefs under the sticker label can hinder reputation building, whereas the laser label enhances consumer learning and strengthens sellers' reputation incentives.

JEL Classification: D22, D83, L11, L14, L15, O10, O12
Key words: Information frictions, quality, consumer learning, firm 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, Christopher Avery, Esther Duflo, Glenn Ellison, Marcel Fafchamps, Hanming Fang, Robert Gibbons, German Sergio Gieczewski, Rema Hana, Gordon Hanson, Seema Jayachandran, Yan Ji, Asim Khwaja, Siyuan Liu, Rocco Macchiavello, Yuhei Miyauchi, Xiaosheng Mu, Ariel Pakes, Harry Di Pei, 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. I thank Xu Yang for the excellent research assistance. 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 issue is exacerbated in markets with information problems, such as food products and pharmaceuticals, where quality is difficult to observe and verify at the point of transaction. In recent years, there has been growing public concern regarding product quality in developing countries.[1] In markets characterized by information frictions and consumer mistrust, firms need a good reputation to succeed. However, many firmsT in developing countries lack a reputation for quality. Empirically, the reputation mechanism does not appear to function effectively in various market settings in developing countries (e.g., Michelson et al. (2021); Bai et al. (2022); Björkman Nyqvist et al. (2022)). The question, then, is: what undermines firms' incentives and ability to build a reputation for quality? Understanding this is crucial to developing solutions that facilitate reputation building and quality provision.

I theoretically and empirically examine this problem in the context of retail watermelon markets in China. I first demonstrate substantial asymmetric information on quality between sellers and buyers, along with a stark absence of a quality premium at baseline. To explain this, I develop a theoretical model that highlights the role of consumer beliefs and costly signaling in shaping sellers' incentives to build a reputation. The theoretical analysis shows that pessimistic consumer beliefs can make reputation building a low-return investment, leading markets to become stuck in a low-quality equilibrium, despite high demand for quality. In such an environment, introducing costly signals can enhance consumer learning and restore sellers' incentives to build a reputation. Motivated by this theoretical framework, I conduct an experiment aimed at inducing quality provision and reputation building by introducing two different signals: a cheap sticker label and an expensive laser-cut label. Consistent with the model, the laser label acts as a costly signal, encouraging sellers to provide higher quality and earn higher sales profits, while the sticker label does not. Finally, leveraging the experimental variation, I estimate an empirical model of consumer learning to recover the evolution of market beliefs. The results show that pessimistic beliefs can hinder reputation building and lead to welfare loss.

The study takes place in local retail markets within a major Chinese city. These markets are typical of many developing countries and serve as the final link in the long supply chain for numerous agricultural products. I focus on fruit stalls selling watermelons, one of the most popular summer fruits, with high consumer demand for quality. However, as with many food products and experience goods, assessing watermelon quality at the point of purchase is difficult (short of cutting it open and tasting it). Fortunately, post-purchase sweetness, an indicator of quality, can be revealed and objectively measured using a sweetness meter. Leveraging this feature, I begin by documenting substantial variation in watermelon quality at baseline, along with a clear absence of a quality premium. In this market, sellers

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

offer undifferentiated piles of watermelons at identical prices, with no seller consistently known for offering sweeter watermelons at a higher price than others. To further explore this puzzle, I conduct a sorting ability test with both sellers and buyers. The results show that while sellers are not perfect at assessing the quality of their watermelons, they are significantly better at it than consumers. This leads to the central question of the paper: given sellers' ability to sort watermelons and the repeated interactions between sellers and buyers, why don't sellers capitalize on their ability to offer a quality premium and build a reputation for quality? Why do we not observe a quality premium or a reputation for quality in these markets?

To examine why the reputation mechanism fails and what it takes to make it work, I first develop a model featuring a discrete-time repeated game between a long-lived seller and an infinite sequence of short-lived consumers. Sellers are of two types: low-cost and high-cost. Only the low-cost sellers can exert effort to sort watermelons in each period, and their effort improves the distribution of watermelon quality. However, due to the information asymmetry, a seller's type and their claim of offering high quality cannot be immediately verified. Consumers can only observe the history of quality and update their beliefs over time. The price a consumer pays in a given period corresponds to their expected quality. The theoretical analysis focuses on the reputation-building incentives of the low-cost type and suggests that consumer beliefs play a crucial role in shaping these incentives.[2] Specifically, pessimistic prior beliefs can make reputation-building a low-return investment. As a result, markets may become stuck in a low-quality, low-reputation equilibrium, even when there is high demand for quality. I then examine the role of introducing a costly signaling device to enhance consumer beliefs. I construct a fully separating equilibrium in which the low-cost type invests in the signal, while the high-cost type does not. Such an equilibrium exists when the signal is sufficiently costly to deter the high-cost type from mimicking the low-cost type. I further discuss the existence of a range of partial pooling equilibria, where a positive fraction of the high-cost type adopts the signal, though this fraction decreases as the signal's cost rises. In sum, the theoretical analysis underscores the role of costly signals in shaping consumer beliefs, with more optimistic beliefs associated with costlier signals. This, in turn, can strengthen sellers' reputation incentives and lead to higher quality provision.

Motivated by the theoretical insights, I design an experiment that introduces different signaling technologies into various markets. The experiment involves 60 sellers operating in 60 local markets in Shijiazhuang, China. In 40 of these markets, I randomly introduce one of two signaling technologies: a cheap sticker label and a more expensive laser-cut label. Consumer surveys suggest that people perceive the laser label as a costly and credible signal of quality. For a cross-randomized subset of sellers, I provide a temporary monetary incentive to invest in higher quality. This incentive aims to subsidize sellers' initial reputation building and inform the history-dependent aspect of reputation formation.

[2] A concurrent theory paper by Pei (2023) examines reputation building under limited observational learning and similarly highlights the importance of consumer learning in sustaining sellers' reputation incentives.

The intervention spans eight weeks, covering the peak watermelon season. On the supply side, I collect high-frequency information on quality, price, and sales through daily field surveys. Quality is measured by randomly sampling watermelons from sellers and testing their sweetness with a sweetness meter. On the demand side, I recruit 675 households from the local markets and collect detailed household fruit purchase and consumption diaries over the intervention period to assess consumer responses.

The experiment yields three main findings, consistent with the theoretical analysis. First, the costly laser labeling induces sellers to offer a genuine quality premium, confirming the presence of reputation incentives. In contrast, the sticker group does not show significantly higher quality or prices than the market average. Second, the incentive treatment successfully motivates sellers to provide higher quality, but this improvement is sustained only for the laser group after the incentive is removed. Third, sellers in the laser group earn 30-40% higher sales profits on average, driven by both higher prices and increased quantities sold, while the sticker group does not outperform the baseline. Overall, the results demonstrate a clear demand for quality and establish the presence of reputation incentives that drive quality provision, particularly among sellers in the laser group. However, in the following season, when the laser labeling is no longer provided, all markets revert to the baseline, lacking quality differentiation. This suggests that small individual sellers may not have sufficient incentive to invest in the expensive technology on their own. The results also indicate a profitable entry opportunity for a larger upstream firm capable of investing in the technology and gradually establishing a reputation.

Additional evidence exploring the dynamics of household purchases and sellers' sales trajectories further highlights the role of consumer beliefs and learning in shaping sellers' returns from establishing their reputation. In the final part of the paper, I incorporate the experimental variation into an empirical model to recover the underlying evolution of beliefs under different signaling technologies. The empirical demand model closely follows the theoretical setup and is estimated using simulated maximum likelihood. Identification relies on the dynamics of household purchase decisions conditioning on reported realizations of consumption experiences. The structural estimates indicate that consumers' prior beliefs are more pessimistic under the sticker labels than under the laser labels. Consequently, establishing a reputation can take a long time, which explains why sellers lack the incentive to provide quality at baseline. In contrast, the laser label enhances prior beliefs and learning, thereby strengthening sellers' incentive to maintain a good reputation. The results confirm the theoretical intuitions and rationalize the experimental findings. Counterfactual analysis quantifies a 21% gain in three-season discounted consumer surplus due to the introduction of the signaling technology.

The study contributes to our broader understanding of consumer learning, firm reputation, and quality provision in markets with information problems. Although the quantitative findings may vary across products and markets, the economic insights are applicable to broader settings. Building a good reputation takes time, and market outcomes can be history-dependent. In markets characterized by pessimistic beliefs and the slow arrival of information signals (e.g., drugs, fertilizers, and food products), low trust and poor quality provision can persist. The issue may be particularly pertinent in developing countries that lack reputable entities and are dominated by small-scale firms. In such environments, it can be difficult for firms to adopt expensive signaling technologies and establish a reputation for quality. Interventions that help markets overcome the information problem and facilitate initial learning and reputation building may yield high returns.

This paper contributes to the empirical literature on consumer learning, firm reputation, and quality provision in markets with information problems. While many studies examine online trading environments, empirical work in the offline world remains relatively sparse (Banerjee and Duflo, 2000; Jin and Leslie, 2009; Macchiavello, 2010; List, 2006; Bardhan et al., 2013; Allen, 2014; Macchiavello and Morjaria, 2015; Startz, 2016; Jensen and Miller, 2018). As noted in Bar-Isaac and Tadelis (2008), a major challenge is that researchers typically do not observe all the information available to buyers or sellers' behavior beyond what buyers observe. This study takes advantage of a field experiment that tracks both sides of the market, demonstrating that consumers' beliefs and the ways they gather information and learn shape sellers' reputation incentives. This aligns with Björkman Nyqvist et al. (2022), which finds that consumer misconceptions hinder the quality provision of anti-malaria drugs in Uganda. Although the contexts differ, the key takeaways are similar. To motivate quality provision, mechanisms that enhance consumer learning or facilitate the entry of large firms may be needed.

The study also relates to the broad literature on firm growth and quality upgrading in development and trade (Verhoogen, 2021). Previous studies have addressed (1) supply-side constraints, including credit access, lack of quality inputs, managerial constraints, and interfirm relationships,[3] and (2) demand-side factors, including access to high-income markets (e.g., Verhoogen (2008); Atkin et al. (2017)). This study highlights another potential barrier to quality upgrading: the information problem and mistrust. Such mistrust, often directed at a broad group level (for example, country-industry), generates an important externality that can hinder individual firms' incentives to upgrade quality (Macchiavello, 2010; Bai et al., 2022).

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 to recover beliefs. Section 7 concludes.

[3] E.g., De Mel et al. (2008); Harrison and Rodríguez-Clare (2009); Kugler and Verhoogen (2012); Banerjee (2013); Bloom et al. (2013); Cai and Szeidl (2017).

2 Setting

Semi-formal, open-air, local markets are one of the most prominent retail venues in developing countries, especially for fresh food products (Grace et al., 2014).[4] Each local market features small-scale retailers operating side by side as shown in Figure A.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 incentive to be strong and to discipline sellers' behavior. However, in recent years, there have been rising complaints about the quality of food products sold in these local markets, in many cases stemming from malpractices of local retailers.[5] Given the repeated interactions and word-of-mouth communication among local consumers, why does the reputation mechanism appear to not function effectively in these markets?

To answer this question, the study focuses on watermelons, one of the most popular products transacted in local markets that represents 35% of household summer fruit consumption in China (as summarized using the baseline household survey in Table 1). Quality of a watermelon can be well-captured using a sweetness meter, shown in Figure A.2, which reports the Brix degree.[6] However, sweetness is hard to detect at the point of transaction. Watermelons are usually sold whole, as cut melons are hard to preserve in hot weather. Sweet and non-sweet watermelons look nearly identical from the outside and are hard for consumers to distinguish (see survey evidence below).

To set up the key empirical puzzle, I document several stylized facts using a combination of baseline surveys and knowledge tests. First, demand for quality appears to be high. To elicit willingness to pay (WTP) for quality, the baseline survey asked households to consider a hypothetical situation wherein two piles of watermelons are sold in local markets: one pile of ordinary quality sells at 1.5 RMB/jin[7]; the other, premium-quality pile sells at a higher price.[8] Figure A.3 plots the empirical distributions of the self-reported WTP for the premium pile. The average quality premium is 28% (1.92 RMB/jin versus 1.5 RMB/jin), and the WTP for quality is higher among households with higher income.

However, despite the seemingly high demand for quality, there is a stark absence of a quality premium. Sellers in a market all sell one undifferentiated pile of watermelons at the same price. The underlying quality, however, varies considerably across watermelons. To document the quality varia-

[4] There is typically an annual fixed fee for operating inside the market. Other than that, these markets are subject to minimal government regulations. Sellers do not need to formally register their business and pay taxes.

[5] Examples include formalin-laced tofu, bean cakes, and rice noodles, water-injected pork and poultry, fossil-adulterated flour, etc. See this 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`. If quality is defined more broadly as value for money, cheating on quantity is ubiquitous in these markets.

[6] A blind tasting test with 210 consumers shows that sweetness strongly correlates with taste: among 210 consumers who were asked to compare two watermelons of high and low sweetness measures, 97% preferred the sweeter one.

[7] 1 jin $\approx$ 1.1 pounds. The rest of the paper uses jin as the pricing unit.

[8] 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.

tion, I randomly sampled 10 watermelons from each of 30 sellers in 30 different markets, representing half of the experimental sample. Panel A of Figure 1 plots the histogram and the cumulative sweetness distribution of the 300 randomly picked watermelons. Sweetness is measured using a sweetness meter. To interpret the scale, a difference of 0.5 matters significantly for taste: Brix degree above 10.5 is considered sweet, and one below 9 tastes plain. Overall, 30% of the 300 watermelons score 9 or below; 43% score between 9 and 10.5; and 27% score 10.5 and above. Notably, 70% of the variation is explained within-seller, suggesting that sweetness varies tremendously within a single batch at a given stall.

Next, to dive into the potential explanations for the lack of quality premium or reputation for quality, I investigate whether sellers have the information about quality and the ability to control the quality of their watermelons, and whether such information and ability are *asymmetric* between sellers and consumers. Retailers are not growers themselves, and they procure their products from the wholesale market in the city, where watermelons of different sweetness levels are not differentiated. Nonetheless, anecdotally, it is well known that local retailers have some ability to assess sweetness through inspections of less obvious observables, such as the skin color, knocking sound, vine curliness, etc. These skills require considerable experiences and are difficult for consumers to acquire.

To formally establish this, I conducted a sorting ability test with the 30 sellers mentioned above and 150 local consumers. Each seller was asked to sort the 10 randomly picked watermelons into two quality piles: one for high quality and one for low quality. I repeated the same sorting test with 5 randomly chosen local consumers in each market. Details are provided in Appendix B.1. Once again, Panel A of Figure 1 shows the distribution of sweetness of all 300 watermelons. Panel B compares and contrasts the sweetness distribution of the high-quality pile sorted by sellers and consumers. The dark gray line plots the distribution of the high-quality pile sorted by the sellers, which is statistically higher than the quality of the unsorted pool (the black line). There also appears to be some heterogeneity in sorting ability among sellers, which I explore further in the empirical analysis in Section 5.4. On the contrary, consumers were not able to assess quality: the light gray line plots the distribution of the high-quality pile sorted by consumers, which is not distinguishable from the unsorted distribution.

The results establish asymmetric information between sellers and consumers and demonstrate that sellers possess some ability, although not perfect, to control the quality of their watermelons.[9] This leads to the central question in the paper: considering all the conditions that appear conducive to reputation building - including high consumer demand for quality, sellers' ability to control quality, and long-term repeated interactions between buyers and sellers - why is it that we do not see a quality premium or a reputation for quality emerging in these markets? It is worth highlighting that a good reputation for quality in this setting means reliably or consistently providing high quality. Give the large amount of variation in quality we see in these markets, having one lucky draw of a sweet watermelon

[9] 5.3% (17.5%) of the high-quality pile sorted by sellers have sweetness below 9 (10). On average, sellers spent 10 seconds per watermelon during the sorting test.

does not indicate that a seller is reliable and consistently offers high quality. While the sweetness of a single watermelon can be immediately discovered upon purchase, inferring *consistency* is much harder. When households were asked in the baseline survey whether they think any seller in the local market consistently provides better watermelons than others, 98% answered "No".[10]

To shed light on this puzzle, I next develop a theoretical model that explains why the reputation mechanism can fail and discuss potential approaches to encourage reputation building.

3 Model

3.1 Setup

Consider a long-run seller with a discount factor δ and an infinite sequence of buyers arriving one at a time in each period. Time is discrete and indexed by $t = 0, 1, 2, \ldots$. In each period, the seller chooses whether to sort her watermelons. There are two types of sellers: low-cost θ_L (skilled) and high-cost θ_H (unskilled), such that it is never profitable for the high-cost type to expend the effort to sort. Suppose that under ordinary conditions without sorting, the probability of a sweet watermelon is $0 < \underline{\gamma} < 1$. A low-cost type can increase the probability of a sweet watermelon to $\overline{\gamma} > \underline{\gamma}$ in any period through sorting, which incurs an effort cost c.

Assume that buyers' valuation of a sweet watermelon is normalized to 1 and valuation for a non-sweet watermelon is normalized to 0. However, due to the information problem, buyers cannot observe the seller's type nor the effort choice in any given period t, but can observe the history of quality draws (sweet or non-sweet) up to t. The price in a given period equals the consumers' expected valuation, that is, the expected probability of drawing a sweet watermelon.

To complete the setup, assume that the cost of offering a non-sorted watermelon is $\underline{\gamma}$ such that the profit of a seller who is known not to exert effort is 0. For a low-cost seller who is known to exert effort to sort, the resulting expected value of her watermelon is $\overline{\gamma}$. Assume $(\overline{\gamma} - \underline{\gamma}) - c > 0$ such that it is profitable for the low-cost type to sort without any information problem.

3.2 Return of Building Reputation

The analysis focuses on the incentive of the low-cost type to build a reputation by exerting effort to sort her watermelons. I begin by describing a non-equilibrium behavioral rule of learning and purchase for buyers, and examine what factors might affect the seller's reputation incentive. The analysis aims to provide several potential explanations for why the reputation mechanism fails to induce quality provision at baseline. The next subsection delves into a full equilibrium analysis, and

[10] 12% of households reported that they usually go to the same seller to buy watermelons, while the majority switch among sellers in their local markets. To quote some, "buying a watermelon is like buying a lottery ticket; sometimes you get a good draw and sometimes you get a bad draw, regardless of whom you go to."

shows that the learning and purchasing behavior outlined here can be supported in an equilibrium with the introduction of a costly signaling device.

Imagine a world in which buyers switch between two regimes: Good and Bad. In the Good regime, buyers expect a low-cost seller to exert effort, while in the Bad regime, they expect the low-cost seller to not exert effort. The high-cost seller is never expected to exert effort. Observing a sweet watermelon in the Good regime preserves the Good regime for another period. Observing a non-sweet watermelon in the Good regime triggers a switch to the Bad regime for the next period. After each period in the Bad regime, there is a stationary probability, denoted by r, of transitioning back to the Good regime. Note that since $\overline{\gamma} < 1$, there is a positive probability of switching from the Good regime to the Bad regime even if the seller has exerted effort to sort.[11]

Suppose the game starts from the Good regime, and buyers' prior belief attaches probability λ^0 to the low-cost type and $1 - \lambda^0$ to the high-cost type. Upon observing the entire history of quality draws in the past, buyers update their belief about the seller's type in a Bayesian manner. Given this, what determines posterior beliefs at the beginning of period t is the number of sweet and non-sweet draws, denoted as m and n, for all past periods in the Good regime. Let $\tilde{\lambda}(m, n)$ denote buyers' posterior belief that the seller is a low-cost type.

Belief updating follows the equations below:

$$\begin{aligned}
&\tilde{\lambda}(0,0) = \lambda^0 \\
&\tilde{\lambda}(m+1,n) = \frac{\overline{\gamma}\tilde{\lambda}(m,n)}{\overline{\gamma}\tilde{\lambda}(m,n) + \underline{\gamma}(1-\tilde{\lambda}(m,n))}, \quad \tilde{\lambda}(m,n+1) = \frac{(1-\overline{\gamma})\tilde{\lambda}(m,n)}{(1-\overline{\gamma})\tilde{\lambda}(m,n) + (1-\underline{\gamma})(1-\tilde{\lambda}(m,n))}
\end{aligned} \tag{1}$$

The seller's reputation is captured by buyers' beliefs $\tilde{\lambda}(m, n)$, which evolves based on the realization of quality draws, with probabilities governed by the (low-cost) seller's choice of effort. Given beliefs, buyers' expected value of a watermelon is $\tilde{\lambda}\overline{\gamma} + (1 - \tilde{\lambda})\underline{\gamma}$, if the period is in the Good regime. For any period in the Bad regime, the expected value for a watermelon is $\underline{\gamma}$ since no type is expected to exert effort. Market price in any given period is determined by the expected value.

Let us now examine the low-cost seller's incentive to build a reputation given the buyers' behavior as described above. Given the belief updating process, the seller would not have an incentive to exert effort to sort in the Bad regime. The question is whether the return of building reputation is strong enough to incentivize the low-cost seller to sort in the Good regime. Let $V_G(m, n)$ denote the value function of a low-cost seller at the start of a period in the Good regime with m and n draws of sweet and non-sweet watermelons in past Good regimes. Similarly, define $V_B(m, n)$ to be the value function

[11] As discussed later in Section 3.3, this is needed to sustain effort in equilibrium as types are revealed, similar to the punishment on the equilibrium path in Green and Porter (1984).

starting from a period in the Bad regime. We can write:

$$V_G^+(m,n) = \tilde{\lambda}(m,n)(\overline{\gamma}-\underline{\gamma}) - c + \delta\overline{\gamma}V_G^+(m+1,n) + \delta(1-\overline{\gamma})V_B^+(m,n+1) \quad (2)$$

$$V_B^+(m,n) = \delta r V_G^+(m,n) + \delta(1-r)V_B^+(m,n) \quad (3)$$

where V_G^+ and V_B^+ indicate the non-negative values of the original value functions, obtained by taking the maximum of the RHS and 0. Intuitively, given any state variables (m,n), the seller always has the option to not exert effort and earn a payoff of 0.

Substituting (3) into (2), we get:

$$V_G^+(m,n) = \tilde{\lambda}(m,n)(\overline{\gamma}-\underline{\gamma}) - c + \delta\overline{\gamma}V_G^+(m+1,n) + \frac{\delta^2(1-\overline{\gamma})r}{1-\delta+\delta r}V_G^+(m,n+1) \quad (4)$$

Figure 2 simulates the value function for different parameter values.[12] To illustrate the main economic forces, the plots vary n for a fixed value of m. When $V_G(m,n)$ hits 0, the incentive for a low-cost seller to build reputation vanishes. Panels (a) show that the seller's return from building reputation depends on the cost of providing quality relative to the buyer's valuation. This is true in markets with and without information problems. Panels (b) and (c) highlight the forces due to the information problem. Panel (b) shows that a lower $\overline{\gamma}$ (holding $\underline{\gamma}$ fixed) reduces the return of building reputation. Intuitively, a lower value of $\overline{\gamma}$ represents worse quality control, which not only reduces the expected value of sorting but also increases the noise in the learning process and reduces the likelihood of staying in the Good regime, thereby slowing down the reputation-building process. Panel (c) shows that when the prior probability attached to the low-cost type λ_0 is low, reputation-building becomes harder (i.e., V_G hits 0 for smaller n). Finally, Panel (d) shows that a smaller r decreases the probability of staying in the Good regime and decreases the return of building reputation.

To further shed light on the role of buyers' beliefs, Panels (e) and (f) of Figure 2 plot the value function against buyers' posterior beliefs $\tilde{\lambda}(m,n)$, and the latter against m and n. Reputation building can be history-dependent: a greater number of bad draws in the past makes reputation building more difficult; the market may be stuck in a no-reputation state with pessimistic initial beliefs. On the other hand, good histories, with more optimistic beliefs, enhance the return of building reputation.

3.3 Introducing a Costly Signal

While the analysis in the above subsection is only suggestive as it does not specify a full equilibrium, the discussion highlights the role of buyers' beliefs in affecting the seller's reputation incentive (i.e., the incentive of the low-cost type to exert effort). Motivated by the discussion, I now consider the

[12] The mathematical proof for the comparative statics is provided in the online appendix: https://drive.google.com/file/d/13vz65Fp5rBgrNLqudqtkmJHbu55xYIjw/view?usp=share_link.

introduction of a costly signaling device and derive the equilibrium implications. In particular, the analysis examines under what conditions such a signal can induce effort to sort for the low-cost type and achieve separation between types.

Consider a pure-strategy separating equilibrium that satisfies the following conditions:

1. Low-cost sellers invest in the signal by incurring a one-time upfront cost M and exert effort to sort in every period in the Good regime, but do not exert effort in the Bad regime.

2. High-cost sellers do not invest in the signal and do not exert effort in any period.

3. Buyers attach probability 1 to the low-cost type upon observing the signal, and 0 otherwise.

4. Observation of a bad watermelon from a seller who has signaled triggers the Bad regime, with a probability r of returning to the Good regime in the following period. Buyers expect the seller to not exert effort in the Bad regime. In the Good regime, buyers expect the seller to exert effort if and only if the seller invests in the signal.

In such an equilibrium,the low-cost sellers manage to distinguish themselves from the high-cost sellers via signaling. Types and actions are fully revealed, and buyers pay the expected value of a watermelon in a given period. In the Good regime, buyers pay $\overline{\gamma}$ to a seller who has signaled, and pay $\underline{\gamma}$ to a seller who has not signaled. In the Bad regime, buyers always pay $\underline{\gamma}$.

The payoff functions for a low-cost seller who has signaled follow Equations (2) and (3) but without belief updating, since the prior probability attached to the low type upon observing the signal is 1:

$$V_G = (\overline{\gamma} - \underline{\gamma}) - c + \delta\overline{\gamma}V_G + \delta(1-\overline{\gamma})V_B \quad (5)$$

$$V_B = \delta r V_G + \delta(1-r)V_B \quad (6)$$

Re-arranging, we get:

$$V_G = \frac{[(\overline{\gamma} - \underline{\gamma}) - c](1-\delta+\delta r)}{(1-\delta)[1+\delta(r-\overline{\gamma})]} \quad (7)$$

Two conditions need to be satisfied for such a pure-strategy separating equilibrium to exist: (1) the low-cost sellers have no incentive to cheat by not exerting effort in the Good regime; (2) the high-cost sellers do not mimic the low-cost type by investing in the signal.

(1) Incentive for effort: First, consider the incentive for a low-cost type (who has signaled) to deviate once by not exerting effort in the Good regime. Such deviation saves cost c but increases the probability of an immediate transition to the Bad regime from $(1-\overline{\gamma})$ to $(1-\underline{\gamma})$. Such a deviation is profitable if and only if $\delta(\overline{\gamma} - \underline{\gamma})(V_G - V_B) < c$. Substituting the values of V_G and V_B from Equations (6) and (7),

we can derive an upper-bound condition on r to prevent cheating and sustain effort:

$$r \leq \frac{\delta(\overline{\gamma} - \underline{\gamma} + \underline{\gamma}c) - c}{\delta c} \equiv r^* \tag{8}$$

Intuitively, the punishment phase triggered by a non-sweet watermelon needs to be long enough (recovery rate small enough) to deter cheating and sustain high effort.

(2) Incentive for signaling: Next, consider the incentive for a high-cost seller to deviate and pay for the signal to mimic the low-cost type. Since high-cost sellers have no ability to exert effort to improve the mix of watermelons, they would act like low-cost sellers who cheat in every period. Define the expected payoffs for a high-cost seller who signals but never exerts effort as W_G and W_B. We have:

$$W_G = (\overline{\gamma} - \underline{\gamma}) + \delta\underline{\gamma}W_G + \delta(1-\underline{\gamma})W_B \tag{9}$$

$$W_B = \delta r W_G + \delta(1-r)W_B \tag{10}$$

Substituting W_B from Equation (10) into Equation (9) and rearranging, we get:

$$W_G = \frac{(1-\delta+\delta r)}{(1-\delta)[1+\delta(r-\underline{\gamma})]} \tag{11}$$

To prevent the high-cost type from mimicking the low-cost type, a necessary condition is to ensure the low-cost type gains more from signaling and exerting effort than the high-cost type from signaling and cheating, i.e., $V_G \geq W_G$. Comparing Equations (7) and (11), we have

$$V_G \geq W_G \quad \textit{iff} \quad r \leq \frac{\delta(\overline{\gamma} - \underline{\gamma} + \underline{\gamma}c) - c}{\delta c} \equiv r^* \tag{12}$$

This is exactly the same condition identified above to sustain effort incentive for the low-cost type.[13]

Finally, to ensure the low-cost type is willing to invest in the signaling technology at the outset of the game but the high-cost type is not, we need $V_G \geq M \geq W_G$.

We are now ready to derive the conditions on M, the cost of the signal, that can support a pure-strategy separating equilibrium satisfying conditions 1-4. First, note that both V_G and W_G increases in r (Equations (7) and (11)). Equations (8) and (12) give the upper bound on r. Define $\overline{M}$ to be the largest value of M such that the low-cost type is willing to invest and exert effort:

$$\overline{M} = V_G(r^*) = W_G(r^*) \tag{13}$$

The second equality follows from Equation (12) that at r^*, the high-cost type gets the same expected

[13]The fact that the two thresholds are the same is an implication of the one-shot-deviation principle: if there is no incentive to deviate once and then revert to the original strategy, then there is no incentive to deviate at every opportunity.

payoff from signaling as the low-cost type.

Next, to derive the lower bound on M, note that the smallest value of r is 0 (corresponding to a grim trigger). However, at $r = 0$, the high-cost seller would still make positive profits if he can costlessly mimic the low-cost seller. In particular, the seller will get a flow payoff of $\overline{\gamma} - \underline{\gamma}$ for some number of periods until a bad watermelon is detected. The expected payoff can be calculated as $W'_H = \overline{\gamma} - \underline{\gamma} + \delta\underline{\gamma}W'_H$, or $W'_H = \frac{\overline{\gamma}-\underline{\gamma}}{1-\delta\underline{\gamma}}$. Define $\underline{M} \equiv \frac{\overline{\gamma}-\underline{\gamma}}{1-\delta\underline{\gamma}}$. We can now state the following result:

Proposition 1: A pure-strategy separating equilibrium satisfying Conditions 1-4 exists for $\underline{M} \leq M \leq \overline{M}$, where $\underline{M} = \frac{\overline{\gamma}-\underline{\gamma}}{1-\delta\underline{\gamma}}$ and $\overline{M} = V_G(r^*)$, where $r^* = \frac{\delta(\overline{\gamma}-\underline{\gamma}+\underline{\gamma}c)-c}{\delta c}$.

Intuitively, when the signaling cost is too low, $M < \underline{M}$, even the grim-trigger punishment (with 0 recovery rate) yields positive profit for the high-cost seller and therefore is insufficient to deter the high-cost seller from mimicking the low-cost type. On the other hand, if the signaling cost is too high, $M < \overline{M}$, it is not profitable for the low-cost type to invest in the signal given the need to provide the incentive to exert effort. In other words, at the highest r^* needed to sustain effort incentive, the expected payoff upon signaling is not large enough to recoup the cost of signaling if $M > \overline{M}$. For $\underline{M} < M < \overline{M}$, there exists a range of r that can support a pure-strategy separating equilibrium.[14]

3.4 Discussion

Proposition 1 highlights that the signaling cost needs to be high enough to be successful at separating the two types. Specifically, the cost needs to be higher than what the high-cost type can gain by mimicking the low-cost type. When $M < \underline{M}$, no full separating equilibrium exists. One may conjecture a partial pooling equilibrium where the low-cost type adopts the signal, and the high-cost type plays a mixed strategy, adopting the signal with a positive probability between 0 and 1. Such an equilibrium requires the high-cost type to be indifferent between signaling and not signaling. Since the market price upon signaling decreases with the fraction of high-cost types that adopt the signal, a lower signaling cost would admit a greater fraction of high-cost types to invest in the signal to satisfy the indifference condition. That is, there will be a greater proportion of high-cost types adopting the signal in a partial pooling equilibrium as the cost of acquiring the signal decreases, resulting in a lower expected payoff and a longer process of building reputation for the low-cost type.[15]

To summarize, the model captures the key features of the watermelon retail markets described in

[14] Panel (a) of Figure A.4 plots V_G and W_G against different values of r, ranging from 0 to 1. r^* is given by the intersection of the two lines; $\underline{M}$ and $\overline{M}$ are determined accordingly.

[15] Panel (b) of Figure A.4 plots the most optimistic beliefs (highest λ^0) buyers can hold upon seeing a costly signal against the cost of the signal for a given set of parameter values. Following Equations (2) and (3), we can derive similar value functions for the high-cost type (upon signaling):

$$
\begin{aligned}
W_G^+(m,n) &= \tilde{\lambda}(m,n)(\overline{\gamma}-\underline{\gamma}) + \delta\underline{\gamma}W_G^+(m+1,n) + \delta(1-\underline{\gamma})W_B^+(m,n+1) \\
W_B^+(m,n) &= \delta r W_G^+(m,n) + \delta(1-r)W_B^+(m,n)
\end{aligned}
$$

Section 2. The discussion highlights a number of potential explanations for the lack of incentive to provide quality and build reputation at baseline. First, the cost of providing quality may be high, relative to consumers' valuation and demand for quality. Second, imperfect quality control slows the learning process and the rate of reputation building. Third, adversarial beliefs can make reputation building a low-return investment. While the different channels act jointly in determining the market outcomes, their implications are very different. If the lack of quality provision is driven mainly by a low WTP relative to cost, then markets may organically evolve to provide higher quality as countries develop and technologies of providing quality improves (as c decreases or $\overline{\gamma}$ increases). However, if the information problem is the main barrier, then introducing effective signaling technologies may yield high return. Through the lens of the equilibrium analysis discussed above, such a signal can shift consumers' beliefs, with more optimistic beliefs associated with costlier signals, which in turn enhances the return of providing quality and building reputation.

4 Experimental Design and Data Collection

Motivated by the theoretical analysis, this section presents an experiment aimed at inducing quality provision and reputation building through two interventions: (1) introducing costly signals and (2) subsidizing the initial cost of quality provision. I first describe the two signaling technologies and the experiment design, and then connect the experiment to the model.

4.1 Two Signaling Technologies: Sticker vs. Laser

Empirically, the effectiveness of different signaling technologies in encouraging reputation building may vary depending on the specific context.[16] The experiment examines two potential signaling technologies: a sticker label of "premium watermelons" ("*Jing Pin Xi Gua*" in Chinese Pinyin) and a laser-cut label with the same words. The key distinction between these two technologies is their costs. Stickers are cheap to print and can be easily pasted onto watermelons, while laser-cut labels require an ex-

Substituting $W_B^+(m,n)$ into $W_G^+(m,n)$, we get

$$W_G^+(m,n) = \tilde{\lambda}(m,n)(\overline{\gamma} - \underline{\gamma}) + \delta\underline{\gamma} W_G^+(m+1,n) + \frac{\delta^2(1-\underline{\gamma})r}{1-\delta+\delta r} W_G^+(m,n+1)$$

It is easy to show that $W_G^+(m,n)$ increases with r and decreases with λ^0. Therefore, the most optimistic belief, the highest λ^0, is given by the indifference condition at the smallest value of r, $r = 0$. That is, $W_G^+(0,0;\lambda^0, r=0) = M$. The figure shows that the highest λ^0 increases with M. At $M = \frac{\overline{\gamma}-\underline{\gamma}}{1-\delta\underline{\gamma}}$, the most optimistic belief converges to 1, as in a full separating equilibrium.

[16] Various signaling technologies have been explored in different market settings, including charging higher prices, utilizing fancy packaging, or implementing various certified labels. Previous literature suggests that in developing countries with prevalent counterfeiting activities, firms often invest in costly branding technologies to protect authentic products and preserve the quality premium (Qian, 2008).

pensive laser machine (approximately $8,000 USD) for engraving.[17] As discussed in Section 3.4, laser labeling involves a larger upfront investment and could potentially induce more favorable initial beliefs.

To verify this, a pre-intervention survey was conducted with 300 consumers in the city of Shijiazhuang in December 2013 (distinct from the experimental sample). The survey asked consumers about their willingness to pay (WTP) for watermelons sold with different "premium-quality" labels, without providing any other information about the underlying quality. Appendix B.2 provides details about the survey questionnaire. The results of the survey showed that, on average, consumers' WTP for sticker-labeled watermelons was low, with only a 4.5% premium compared to ordinary watermelons, and the difference was not statistically significant. On the other hand, consumers' WTP for laser-labeled watermelons was 23% higher, and this difference was statistically significant at the 5% level. Interestingly, when consumers were asked about the reason for their higher WTP for laser-labeled watermelons, 78% of them responded that they regarded laser labeling as a more credible signal of quality because it is expensive and difficult to forge by low-quality sellers. In contrast, stickers can be cheaply made, and as a result, the quality signal they provide can be diluted by the presence of low-quality products. These responses highlight the generalized mistrust among Chinese consumers of sticker-labeled food products, partly due to past counterfeiting activities related to various quality certificates issued in the sticker form.[18]

Mapping these qualitative accounts to the model in Section 3, the expensive laser label serves as a more credible and effective signal by deterring the entry of low-quality sellers (the high-cost type). Therefore, it has the potential to boost buyers' initial beliefs and demand. This, in turn, may strengthen sellers' reputation incentives and potentially induce quality provision. On the other hand, the sticker label is likely to be an ineffective quality signal since it can be cheaply produced. As a result, if beliefs are pessimistic, sellers may lack the incentive to provide quality and build reputation.

4.2 Experimental Design and Timeline

The experiment was conducted in Shijiazhuang, China.[19] The city has over 800 gated communities and more than 200 local markets. Randomization was carried out at the market level. A total of 60 sellers located in 60 different markets were recruited to participate in the study, following an initial screening

[17] The laser technology was invented by a large agricultural company in China, Hebei Shuangxing Seed Co., Ltd., to brand the company's high-quality watermelons. It was put into use a year after the experiment in multiple cities in China. The study was done in collaboration with the company, which provided the research team with the laser machines.

[18] The credibility of various certified labels in the sticker form, including "pollution-free," "green," and "organic," has been undermined by widespread forgery issues. An article from *Huanqiu* and a CCTV report shed light on the prevalence of fake certifications in China, emphasizing the need for credible and trustworthy signals to ensure consumer confidence in product quality: `http://finance.huanqiu.com/pictures/2011-10/2127997.html` and a CCTV report about fake certification of green food and organic food products in China at `http://m.news.cntv.cn/2014/09/14/ARTI1410670045348762.shtml`.

[19] The city has a total urban area of 154.2 square miles and a population of 2,861,784 people, with an urban density of approximately 19,000 people per square mile.

procedure to minimize heterogeneity in the study sample for both power and logistical purposes. Details of the screening process and selection criteria are presented in Appendix B.3.

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. Sellers were free to choose the quality, price, and quantity of each pile.[20]

Sellers were randomized into 6 groups:

Labeling treatments. Sellers were randomly assigned to one of three labeling groups: laser, sticker, and label-free. Each morning, surveyors visited the sellers' stores and provided a free labeling service. For the laser group, the surveyors used a laser-engraving machine to laser-cut the words "premium watermelon" onto the watermelons in the premium pile, which were sorted by the sellers themselves. For the sticker group, surveyors pasted a sticker with the same words onto the watermelons. The label-free group did not receive any labeling service.

It's important to note that labeling was only done for watermelons in the premium pile that were picked by the sellers themselves. Watermelons in the normal pile were left unlabeled. Figure 3 shows pictures of the labeling treatments. Initially, most sellers had two piles of watermelons, but some reverted to non-differentiation after some time. For those sellers, the labeling service was withdrawn as there was no longer a premium pile.

A cross-randomized incentive treatment. Within each labeling group, half of the sellers were randomly assigned to an incentive treatment aimed at subsidizing the initial cost of reputation building to encourage high quality provision. It involved unannounced quality checks twice per week. During each check, surveyors randomly selected one watermelon from the premium pile and one from the normal pile. The sweetness of both watermelons was measured using a sweetness meter. For sellers in the incentive group, if the sweetness of the selected watermelon from the premium pile was at least 10.5 in both checks, they received a monetary reward of 100 RMB at the end of the week (equivalent to average daily sales profits). To minimize concerns about collusion between sellers and surveyors, the surveyors were rotated across markets on a weekly basis. Sellers in the non-incentive group also received the same quality checks but were not offered a reward.

The incentive was removed in week 6 after the intervention began, and this removal was unanticipated by the sellers. In 71% of the eligible seller-week observations, the quality requirement for the incentive group was met, and the monetary reward was issued.[21]

Summary. In total, there were 6 distinct treatment arms. Randomization was stratified based on housing prices, which served as a proxy for local income and potential demand for quality. Figure A.5

[20] Sellers participating in the study received a fixed payment of 100 RMB per week. This compensation was mainly provided to compensate for the time they spent recording their daily sales, as described in Section 4.3.

[21] Eligible seller-week observations included all watermelons randomly picked from sellers in the incentive group during the first 6 weeks, when the sellers sorted watermelons into two piles, one of which was the premium pile.

shows a map of the 60 sellers, marked by treatment groups. It is worth noting that these markets are geographically segregated, with the distance between the two closest markets being approximately 1 kilometer. As watermelon transactions are highly localized, spillover effects across markets should be minimal.[22] That said, there could be spillover effects to or strategic responses from the non-sample sellers operating in the same 60 markets.[23] Data on the other sellers' pricing and differentiation behavior were collected to examine any potential spillover effect.

Timeline. Figure 4 describes the timeline of the experiment. The intervention was rolled out 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. This allowed for the examination of 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' stalls, and two follow-up surveys were conducted to examine longer-term outcomes.

4.3 Data

Baseline surveys. Table 1 summarizes the baseline characteristics of the 60 sellers, the local markets and the 675 households in the experimental sample. The majority of sellers engage in year-round fruit sales and have no plans to relocate. The median household consumes 1 watermelon per week during the summer, with 75.6% of households listing the local market as their main source of purchases.

Supply side: quality, prices, and sales. Enumerators collected daily retail prices for both the sample sellers and other sellers in the markets, as well as the daily wholesale price. Quality data were collected through the biweekly random quality checks. In addition, sellers were asked to record their daily sales of watermelons by quality category.[24] Throughout the intervention, a total of 49,253 transaction records were collected. On average, sellers sold 257 jin ($\approx$ 340 pounds) of watermelons per day, and the average daily sales profits were 103 RMB. For the empirical analyses, sales profits are calculated by multiplying sales quantity with the difference between retail and wholesale prices. This does not take into account any transportation/storage costs or effort costs of sourcing higher quality.[25] For most of the analyses, transaction-level sales are aggregated to the seller-day-quality category level.

Demand side: household panel purchase and consumption experiences. A total of 675

[22] According to the baseline survey, 80% of watermelons are bought from a given household's local market, while the remaining 20% are purchased from nearby supermarkets rather than other local markets. During the experiment, there were few instances of consumers switching between markets, such as from a label-free market to a laser market.

[23] On average, each market contained 3 fruit sellers (Table 1), with only 1 included in the study sample.

[24] Panel A of Figure A.6 provides an example of a recording sheet, and Panel B presents an example of the household recording sheet.

[25] Specifically, sales profits = 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, accounting consumer bargaining. The results are quantitatively robust.

households in 27 communities, evenly distributed across the treatment groups, were recruited to record their entire summer fruit purchase and consumption experiences. For each purchase, households were asked to record the date and 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 labels on it. Additionally, households were asked to rate their consumption experience on a scale from 1 to 5, where a higher number indicates a higher level of satisfaction. This allows us to observe individual experiences and examine belief updating. In total, there are 15,292 purchase records, with 30.8% for watermelons. The median number of watermelons consumed per week is 1, and the mean is 1.15 with a standard deviation of 1.06. These numbers match the baseline summary statistics in Table 1. Appendix B.5 discusses additional issues and cleaning of the household data. Households with a high number of missing transaction records are excluded from the main analysis. The final analysis sample consists of 573 households with more than 4,300 watermelon purchase records.

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

Details of the sampling and recruiting procedure, data collection, and issues with cleaning the seller and household recording data are discussed in detail in Appendix B.4 and Appendix B.5. Balance checks on market, seller, and household baseline characteristics are provided in Tables A.1 to A.3.

4.4 Connecting the Experiment to the Theory

Before presenting the experimental results, it is important to highlight several conceptual points that connect the experiment to the model in Section 3. These points help to understand the implications of the experiment and interpretation of the experimental findings.

First, the labeling treatments (both laser and sticker) were provided to sellers at zero cost and could be used at the seller's discretion. This design was necessary to isolate pure *reputation* incentives from third-party quality enforcement or certification. However, while the labeling was provided for free to sellers, this information was not revealed to consumers.[26] Therefore, consumers may still have

[26] In practice, the labeling services were carried out very early in the morning before consumers arrived, and it was not in sellers' interests to disclose the information about the experiment to consumers.

perceived the laser labels as a positive quality signal based on the equilibrium analysis in Section 3.3. From the seller's perspective, more optimistic prior beliefs enhance the return of building reputation, and that can be sufficient for the treatment to work. That said, it would be difficult to map the observed actions of sellers under the experiment exactly to the equilibrium analysis in Section 3.3. The experiment essentially aims to nudge sellers to behave off the equilibrium path to shed light on why the market is stuck in the bad equilibrium in the first place.

Relatedly, it is important to distinguish between short-run and long-run incentives and outcomes. In the short run, if the labeling treatments induce sellers to provide quality during the intervention, that itself establishes that reputation incentives are present and can motivate quality provision. However, an open question is whether sellers will continue to provide quality when the labeling technologies are no longer provided for free. The answer depends on whether any reputation developed under the treatments is associated with the individual retailers. Based on the experimental design, it is possible that consumers may perceive the labeled watermelons as coming from some upstream suppliers. The model in Section 3 still applies to interpret the demand-side responses under the experiment but leaves open the question of sellers' responses post-intervention. In particular, would sellers have the incentive to invest in the technologies themselves, and if so, why have they not done so at baseline? I address these questions in Section 5.6 after presenting the main experimental findings.

Last but not least, the incentive treatment represents a way of subsidizing sellers' initial reputation building. By doing so, it raises sellers' per-period profits and increases the expected payoff of building reputation, regardless of market beliefs. If such an initial incentive does motivate sellers to provide higher quality, then over time, as consumers try the products and update their beliefs, sellers who received the incentive will essentially be endowed with a higher reputation than those who did not receive the incentive. The market can reach a point where beliefs are favorable enough that sellers who had the incentive will continue to provide quality and maintain their good reputation even after the incentive is removed. This, therefore, provides a further test for the model. The incentive treatment also allows me to compare sales and quality dynamics *within* each labeling treatment group, which helps to address several alternative explanations aside from the learning and reputation mechanism.

5 Experimental Evidence

This section presents the experimental findings. Sections 5.1, 5.2, and 5.3 examine the impacts of the labeling and incentive treatments on sellers' quality provision, pricing, sales, and profits. Section 5.4 provides suggestive evidence of heterogeneity across sellers. Section 5.5 sheds light on the impact of different labeling technologies on household learning and purchasing dynamics. Section 5.6 ties the experimental findings together to explain the lack of reputation and quality provision at baseline. Section 5.7 discusses alternative explanations.

5.1 Effects of the Labeling Treatments on Sellers' Quality Provision

Figure 5 plots the number of sellers who differentiated quality at sale in each treatment group over time. We observe that sellers in the label-free group sharply reverted to non-differentiation after the first two weeks, once quality differentiation was no longer enforced. This behavior is consistent with their baseline practices. In contrast, most sellers in the sticker and laser groups continued to differentiate quality throughout the entire intervention period. The patterns for the non-incentive group (Panel A) and the incentive group (Panel B) exhibit similar trends.

Next, I look at sellers' quality provision conditioning on differentiation at sale, focusing on the sticker and laser groups. Panel A of Table 2 compares the premium pile quality, measured in sweetness, for sellers in the sticker and laser groups. Columns 1 and 2 pool together both the incentive and non-incentive subgroups, while Columns 3 and 4 further restrict the sample to those sellers in the non-incentive groups to isolate the impact of the labeling treatments. Standard errors are clustered at the seller (market) level, which is the unit of randomization. To address concerns over the relatively small sample size and the small number of clusters, I also conduct two small-sample robustness checks using a permutation test (Bloom et al., 2013) and a clustered bootstrap (Cameron et al., 2008). The p-values are reported in the table. On average, sellers in the laser group provide significantly higher quality than sellers in the sticker group. The same pattern holds when using household satisfaction rates as a measure of quality, as shown in Table A.4.

To further investigate sellers' quality provision, I examine how the quality of the premium pile compares to that of the normal pile and the market average. Columns 1 and 2 of Table 2 Panel B show that the average quality of the premium pile is significantly higher than that of the normal pile. However, this difference could be due to either genuine quality improvement in the premium pile or a quality deterioration in the normal pile. To examine these possibilities, I compare the quality difference from the market average. Columns 3 and 4 run the same regression, but with the *quality difference* from the market average as the outcome variable. I use the average sweetness of randomly picked watermelons from sellers in the label-free group after they reverted to non-differentiation as a proxy for average quality. Column 3 shows that sellers in the laser group offered higher quality in the premium pile while keeping the normal pile quality on par with the market average. On the other hand, for the sticker group, the average quality of the normal pile is lower than the market average, and the quality premium for the premium pile is not significantly different from 0, as shown in Column 4 (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 observably bad ones, which they then marked down and sold as a low-end product.

The findings from the laser group highlight the incentives for reputation building. In a one-shot game, sellers might not exert additional effort to provide higher quality and would randomly label some watermelons as "premium" to sell them at a higher price. However, the results show that sellers in the laser group did put more effort into sourcing good watermelons. Qualitative evidence from a follow-up survey supports this, with 85% of sellers in the laser group reporting expending efforts to search for better watermelons in the wholesale market, compared to 65% and 60% in the sticker and label-free groups, respectively. Additionally, sellers in the laser group reported spending more time in the wholesale market sourcing watermelons than the other groups, with an average of 52.5 minutes compared to 43.5 minutes (as shown in Figure A.7).

Overall, these findings align with the theoretical model's prediction that a more expensive and harder-to-fake signal can enhance sellers' reputation incentive and motivate sellers to invest more effort in providing higher quality watermelons.

5.2 Effects of the Labeling Treatments on Prices, Sales and Profits

Table 3 examines the effects of the labeling treatments on sales outcomes. The outcome variables in this analysis are measured at the seller-day level and include log sales profits (in RMB), price premium above the market average price (in RMB/jin), sales quantity (in jin) for each pile, and the total sales quantity.[27] In cases where a seller stops differentiating quality, the unit price for the premium pile is set to be the same as that for the normal pile, and the sales quantity for the premium pile is coded as 0. This allows for a consistent comparison across sellers with varying differentiation behaviors. To account for time-specific aggregate shocks, such as weather conditions, all regression models include day fixed effects. The even columns control for community and seller baseline characteristics.

In Columns 1 and 2 show that, on average, the laser group experiences 30-40% higher sales profits compared to the label-free group. This increase in profits can be attributed to both a higher price (shown in Columns 3 and 4) and a greater sales quantity for the premium pile (seen in Columns 5 and 6). The sales for the normal pile are not significantly different from those of the label-free group. These findings suggest that sellers in the laser group are able to attract more high-end customers without losing sales from the normal pile. It is worth noting that other competitors in the market did not follow the same strategy of quality differentiation, as they were not provided access to the technology used by the sample sellers. There were also no significant strategic pricing responses observed among the other sellers (as shown in Table A.5).

On the other hand, for the sticker group, sales from the premium pile appear to be lower on average than those for the laser group (the p-value of a one-sided test is 0.238) despite having a lower price. Furthermore, the increase in sales from the premium pile (as shown in Columns 5 and 6) is offset by a reduction in sales from the normal pile (as indicated in Columns 9 and 10). As a result, total sales

[27] Here and in all subsequent analyses with prices, I use the listed prices observed by enumerators 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 from the sellers' sales records. The results are similar.

and profits for the sticker group are not significantly different from those of the label-free group, which reverted to non-differentiation. These findings help explain why sellers did not differentiate quality at baseline, even though stickers have long been cheaply available.

5.3 Effects of the Incentive Treatment on Sellers' Quality Provision

Table 4 presents the results of a difference-in-difference regression to examine the quality provision before and after the incentive was removed. Overall, we observe that the incentive treatment led to higher quality provision for both the sticker and laser groups. 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. This suggests that high quality provision was sustained for sellers in the laser group even after the incentive was removed.

On the other hand, for the sellers in the sticker incentive group, there seems to be a decrease in quality provision after the incentive was removed. These results align with the theoretical discussions mentioned earlier: if beliefs are more pessimistic under sticker labeling, then reputation building would take longer. Thus, it is not as clear how much the incentive facilitated initial reputation building during this relatively short intervention. Additional results on the interaction between the labeling and incentive treatments are in Table A.6. Overall, the laser incentive group provides the highest average quality among all the groups, followed by the sticker-incentive and laser non-incentive groups.

5.4 Heterogeneous Treatment Effects among Sellers

The theoretical model posits the existence of different seller types based on their ability or costs of sorting, and these different types of sellers may exhibit different behaviors when provided with the labeling technologies. Specifically, the model suggests that sellers with higher sorting ability or lower costs of sorting will find it more profitable to use the labeling technologies to differentiate their products based on quality. On the other hand, sellers with lower sorting ability or higher costs of sorting may not differentiate their products, even if they are provided with the technology.

To examine such heterogeneity, I utilize the sorting ability test described in Section 2 and measure a seller's ability based on their performance in the test. Specifically, I assess whether a seller made any "clear mistake" at the sorting test. A clear mistake occurs when at least one watermelon sorted to the low pile strictly dominates the quality of one (or more) watermelons sorted to the high pile. Among the 30 sellers who participated in the sorting test, 7 made such clear mistakes. Table A.7 presents the correlation between ability and seller characteristics. It shows that ability is positively correlated with the years of experience in selling watermelons, with male sellers showing better sorting ability compared to female sellers. Interestingly, community characteristics, such as housing price and number of housing units, do not predict ability, suggesting that sellers do not sort into different markets based on their ability. This observation aligns with the fact that quality is not priced at baseline, which provides limited incentives for sellers to choose locations based on their sorting ability.

Table 5 examines the heterogeneity in pricing and quality provision during the intervention based on the ability measured at the sorting test. Given that only 30 sellers participated in the sorting test at baseline and only once, the analysis is suggestive and should be interpreted with caution due to the limited sample size. The results suggest that sellers with higher sorting ability tended to charge a higher price for their premium pile and provide higher quality products. This heterogeneity is particularly noticeable for the laser groups (Column (2) and (4)). Even though the laser machine was provided for free, not all sellers who received it managed to consistently provide higher quality, consistent with the existence of different seller types in the population.

5.5 Effects of the Labeling Treatments on Household Purchasing Dynamics

The treatment effects on the supply side support the interpretation that the more expensive laser labeling technology enhanced consumers' initial beliefs and learning. This, in turn, strengthened the reputation incentive for sellers to provide higher quality products. However, identifying the impacts of different labeling technologies on beliefs is challenging given that beliefs are not directly measured in the data. Section 6 addresses this challenge by estimating a structural model of learning and belief updating, allowing us to recover beliefs through the lens of the model. Here, I present additional reduced-form evidence using the household panel data, focusing on household purchasing dynamics in response to past consumption experiences, which reflects underlying belief updating.

Intuitively, if buyers initially hold pessimistic beliefs about product quality, a credible signaling technology that improves their beliefs would also increase the variance (i.e., reduce stubbornness) in their initial beliefs, leading to faster belief updating.[28] As a result, we would expect buyers to become more responsive to past consumption experiences. To investigate this empirically, I leverage the household panel data, which includes information on both the purchase decisions of households and their self-reported satisfaction ratings for each consumption experience.

Table 6 examines the impact of past experiences on future purchases. The data are aggregated to the household-week level for analysis. The dependent variable is a binary indicator for whether a household purchased any watermelons, either premium or normal, from the treated seller in a given week. It is regressed on two measures of past purchase experiences from the same seller of a particular pile: (1)

[28] I use the theoretical model presented in Section 3 to illustrate this point. In Panel (c) of Figure A.4, I plot $\tilde{\lambda}(m, n)$ against different numbers of sweet draws while keeping the number of non-sweet draws fixed at 0. Each line represents different initial beliefs, achieved by varying λ^0. Since beliefs about types are binary, the variance is given by $\lambda^0(1 - \lambda^0)$. For small values of λ^0 (up to 0.5), the variance increases with λ^0, and it decreases as λ^0 further increases. The variance of beliefs governs the degree of belief updating following sweet (and non-sweet) draws, which subsequently influence future purchasing decisions. An increase in variance indicates that buyers become more responsive to past consumption experiences, as depicted by the steeper slope of the plotted line.

the average lagged satisfaction rating of all past purchases and (2) the percentage of past purchases that received the highest rating of 5. Note that if a household never purchased any watermelons from the seller in the past, these measures are not defined. Therefore, the coefficients are estimated from household-week observations, conditioning on a positive number of purchases prior to a given week.

Panel A presents the results on the purchasing dynamics of the premium pile, separately for households in the laser markets and the sticker markets. Columns 1 and 2 demonstrate that lagged experiences strongly predict repurchasing decisions for households in the laser markets. To interpret the magnitudes, consider 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 has experienced the product). On the other hand, the coefficients are much smaller and statistically insignificant for households in the sticker markets, as shown in Columns 3 and 4. These findings are consistent with discussion in Section 4.1 that beliefs under stickers may be pessimistic. On the other hand, laser labeling improves both the prior mean and prior variance, both of which enhance the speed of reputation building.

Panel B repeats the same analysis for purchases from the normal pile. Since consumers are used to purchasing unlabeled watermelons, each additional experience should not significantly shift their beliefs and influence future purchases. As expected, the coefficients are small and statistically insignificant.[29]

5.6 Resolving the Puzzle and Going Beyond the Experiment

Overall, the experimental findings support the learning mechanism and the role of costly signals in influencing consumers' beliefs and sellers' reputation incentives. Why, then, is there a lack of quality provision and premium at baseline? In other words, would sellers have the incentive to adopt the costly signaling technology themselves, and if so, why had they not done so at baseline?

While the introduction of the expensive laser technology led to a 30-40% increase in sales profits for the laser group, the small market size of each seller, with average sales profits of 4,226 RMB during the intervention (which lasted approximately for one summer season), means that it would take close to 12 years to recoup the fixed cost of the laser machine, not accounting for any effort cost of sorting. In general, a signaling technology needs to incur substantial upfront costs in order to be credible. However, these high fixed costs can present a significant barrier to adoption among small-scale sellers

[29] One caveat is that consumers may hold different criteria for "satisfactory" watermelons purchased under different labeling technologies. If a satisfactory laser-labeled watermelon is in fact better than a satisfactory sticker-labeled watermelon, we might see a stronger relationship between past satisfactory experiences and purchases in the laser markets, which confounds the learning story. To examine this, I take advantage of the incentive treatment and compare the effect of lagged satisfaction on future purchase between the incentive and non-incentive markets. Table 4 shows that the incentive does indeed lead to higher provision of quality for both the laser and sticker groups. Table A.8 shows that the satisfaction-purchase relationship looks similar in both the incentive and non-incentive markets.

in developing countries. Instead, larger upstream wholesalers may be better positioned to invest in such technologies and build a reputation for quality.[30]

Two related questions follow. First, can a third party invest in the technology and rent it out to sellers at a subsidized price? The theoretical discussion in Section 3 highlights that if the premium for sweet watermelons exceeds the price of the labels, even sellers with non-sweet watermelons would buy the labels, rendering the signal worthless. This issue applies to third-party certifications. If a third party with expertise in sorting watermelons were to issue costly certificates for those meeting the premium standard, it could work similar to laser labeling in signaling quality. However, a challenge in many developing countries is the fabrication of quality certificates issued in conventional forms, such as stickers and papers. This raises concerns about the credibility of certifications, as they can be easily forged, diminishing their effectiveness as quality signals in the market.

Another question is whether a one-time intervention is sufficient to induce long-term changes in sellers' behavior and consumer beliefs. One year after the intervention, none of the 57 sellers that could be tracked continued with quality differentiation. This suggests that the good reputation developed with the laser technology may not have been attached to the local retailers. Instead, consumers could have interpreted the laser labeling as an upstream branding. This is plausible since all sellers in the local markets source watermelons from the same wholesale market in the city. Moreover, the cost of the laser machine is prohibitive for small local retailers, as discussed earlier, further supporting the perception of an upstream branding. In this context, consumers would be learning about the laser brand rather than the specific local retailer. Once a seller no longer carries that brand, consumers may no longer believe that the sellers would provide quality, and consequently, there would also be no incentive for the seller to offer high quality.

The discussion above highlights potential challenges of building reputation in supply chains where the goods change hands multiple times (from farmers to traders to wholesalers and to local retailers) and every party along the chain may alter quality. The theoretical framework in Section 3 focuses on a single seller's reputation incentive, but additional complexities may arise in multi-layered supply chains. Future research is needed to understand how reputation incentives can be effectively distributed and maintained along supply chains.

5.7 Alternative Explanations

One alternative explanation for the success of laser labeling is that it is perceived as "cool" and adds direct utility to consumption. However, such an effect alone would not explain the purchasing dynamics discussed in Section 5.5, which goes beyond a static "coolness effect" and supports a learning story.

[30] As seen in subsequent years after the experiment, Hebei Shuangxing Seed, the inventor of the laser technology, successfully sold laser-branded watermelons in multiple cities in China.

To examine this further, I exploit the cross-randomized incentive treatment and compare sellers' sales dynamics within the same labeling treatment group. As we see in Table 4, the incentive treatment led to higher quality provision. Figure A.8 shows that sales performance diverges over time between the laser incentive and non-incentive groups. This divergence aligns with the earlier finding that the former provided higher quality, and over time, higher quality yields higher sales as consumers experience the product and update their perceptions. To quantify the differential sales growth, Table A.9 estimates a linear time model and finds significant positive coefficient for the interaction between the incentive treatment and time for the laser group. Interestingly, we do not see such a dynamic pattern for the sticker groups, consistent with slower belief updating under sticker as discussed in Section 5.5.

It is also important to acknowledge the role of relationships, which commonly exist in these markets (Fafchamps, 2002). For instance, sellers may selectively offer higher quality watermelons to repeat customers. The lack of explicit quality differentiation at baseline would not be a problem if relational contracting perfectly allocates high-quality watermelons to consumers with high valuation for quality. However, if that were the case, we would not expect to see the positive effect on sales for the laser group. To the extent that sellers' preferential treatment may not perfectly align with consumers' willingness to pay for quality, there could still be important welfare loss due to misallocation.

6 An Empirical Model of Consumer Learning and Seller Reputation

The theoretical and experimental findings have highlighted the importance of consumer learning in shaping sellers' reputation incentives. To delve deeper into this process and shed light on the dynamics of consumer beliefs and seller reputation, I extend the theoretical model presented in Section 3 to estimate an empirical model of demand in the watermelon market. This empirical model closely follows the setup of the theoretical model and incorporates several key empirical features of the market.

6.1 Setup and Assumptions

Prior beliefs and belief updating. Following the theoretical setup in Section 3.1, consumers hold a common prior beliefs, λ^0, about the type of sellers when presented with a premium pile of watermelons. These prior beliefs may depend on the specific signaling technology used, denoted as λ_s^0 for the sticker technology and λ_l^0 for the laser technology. The experiment introduces random variation in the signaling technologies among sellers, which helps to identify the difference in prior beliefs. Specifically, households living in different markets face different choice sets: Households in the laser markets, denoted as $M(s)$, are presented with a premium option labeled with the laser technology. Households in the sticker markets, denoted as $M(s)$, are presented with a premium option labeled with the sticker technology. Finally, households in the label-free markets, denoted as $M(ll)$, face a choice set without the premium option.

In each market, consumers do not directly observe the actual quality of the premium pile at the time of the transaction. Instead, they rely on the signal provided by the seller (either the sticker or laser label) and their past consumption experiences to update their beliefs about the seller's type. Belief updating follows Equation 1. In particular, for an individual consumer i in market m, if the last period's consumption experience is good, the consumer stays in the good regime; beliefs in period t is given by $\tilde{\lambda}_i^t = \tilde{\lambda}(\lambda_m^0, m_{i,t}, n_{i,t})$, as specified in Equation 1. λ_m^0 depends on the signaling technology the seller is randomized into; $m_{i,t}$ and $n_{i,t}$ denote the past good and bad experiences of individual i up to period t (in good regimes). However, if the last period's consumption experience is bad, the consumer enters into the bad regime, in which the seller is believed not to exert effort to provide quality. In the following period $t+1$, with probability r, beliefs switch back to the good regime; with probability $1-r$, the bad regime persists.

Purchase. Consumers make purchase decisions in a given period based on their posterior beliefs. To model purchase behavior, I extend the theoretical demand model using a discrete choice logit framework to incorporate richer purchase options as well as an outside option of choosing not to make a purchase.

Consider three purchase choices of watermelons: $j \in \{1, 2, 3\}$, 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 sellers in the market. For the premium pile, given the posterior belief $\tilde{\lambda}_i^t$, the expected quality is $\tilde{\gamma}_{im1t} = \tilde{\lambda}_i^t \overline{\gamma} + (1 - \tilde{\lambda}_i^t)\underline{\gamma}$. For the normal pile and those from other sellers in the market, I assume that consumers do not update their beliefs on these options, and thus, $\tilde{\gamma}_{imjt} = \underline{\gamma}$, for $j = 2, 3$, and all i, m, t. This assumption is motivated by the reduced-form results in Panel B of Table 6, which find no salient patterns of belief updating for the normal pile.

I further enrich the empirical model to account for any direct utility associated with consuming labeled (branded) watermelons, which could vary for laser and sticker labels. Additionally, I consider the possibility that consumers may downgrade their perception of the normal pile if they notice that the same seller also offers a premium pile. Specifically, the expected utility of consumer i for purchasing option $j \in 1, 2, 3$ at time t is given by:

$$
\begin{aligned}
u_{imjt} \quad &= \tilde{\gamma}_{imjt} - \alpha P_{mjt} \\
&+\eta \mathbf{I}(j=1) + \eta_l \mathbf{I}(j=1, m \in M(l)) + \tau_s \mathbf{I}(j=2, m \in M(s)) + \tau_l \mathbf{I}(j=2, m \in M(l)) \\
&+\nu_m + \nu_t + \epsilon_{imjt}
\end{aligned}
$$

where $\tilde{\gamma}_{imjt}$ represents i's posterior mean quality for option j as described earlier. P_{mjt} is the price of option j in market m at time t. The parameter α represents the price coefficient. η captures any time-invariant taste for the premium option, which could include any direct utility associated with consuming labeled (branded) watermelons. η_l allows for different effects for the laser markets. τ_s and τ_l represent the potential spillover effects to the normal pile when sellers also offer a premium pile. ν_m captures market fixed effects, accounting for time-invariant differences across markets. For example, consumers in some markets may consume, on average, more watermelons than those in other markets. ν_t represents time fixed effects, capturing aggregate time trends or shocks that affect all markets. For example, consumers may buy more watermelons on sunny days compared to rainy days. ϵ_{imjt} denotes idiosyncratic random utility shocks realized in each period before the purchasing decision is made. Let V_{imjt} denote the mean utility, excluding the random shock.

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

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

6.2 Estimation and Identification

The model consists of ten structural parameters: $\{\lambda_s^0, \lambda_l^0, \underline{\gamma}, \overline{\gamma}, r, \alpha, \eta, \eta_l, \tau_s, \tau_l\}$, in addition to the vector of markets and time fixed effects, $\{\nu_m\}$ and $\{\nu_t\}$. I first calibrate $\underline{\gamma}$ and $\overline{\gamma}$. For $\underline{\gamma}$, I use the quality sampling data from the label-free group, where 73% of watermelons have sweetness above 10.5 (the threshold used for the incentive). For the baseline estimation, I set $\underline{\gamma} = 0.3$. For $\overline{\gamma}$, which reflects seller's innate ability to sort, I leverage the sorting ability test conducted at baseline. On average, 56.5% watermelons in the premium pile sorted by seller are above 10.5 in sweetness. The median is 0.625 and the 75th percentile is 0.75. I set $\overline{\gamma} = 0.625$ in the estimation; results using alternative values of $\overline{\gamma}$ are qualitatively similar (not shown).

To estimate the remaining parameters, I use the method of simulated maximum likelihood (Train (2009)). I aggregate the household panel purchasing data to the household-week level and merge it with the market-week level average prices.[31] Each purchase experience is associated with a reported satisfaction rating ranging from 1 to 5. I recode the ratings such that 5 represents a satisfactory experience, while $1, 2, 3, 4$ and missing values indicate non-satisfactory experiences. Using this classification, the empirical satisfaction rate among households in the label-free market is close to 30%, aligning well with the 10.5 threshold in sweetness.

The identifying assumption is that the market and time fixed effects fully capture unobserved time-varying shocks that directly affect both prices and demand within each market. With one-period data on market shares, we can identify the time-invariant parameters η, η_l, τ_s, τ_l, market fixed effects,

[31] In some cases, a household may make multiple purchases in a given week. To accommodate this, I apply the Bayesian updating formula multiple times based on all the realized experiences in that week.

and the price coefficient α following standard arguments in the discrete choice literature. Intuitively, these parameters affect the *level* of purchases at time 0. The key parameters of interest, λ_s^0, λ_l^0, and the recovery rate r, are identified from the dynamic repurchasing decisions conditioned on past experiences. Higher values of λ^0 not only lead to higher initial purchases but also faster speed of learning (for $\lambda^0 < 0.5$) and thus a larger increase in repurchasing probability following a positive experience. Similarly, conditioning on beliefs, the repurchasing probability following a negative experience informs the recovery rate r.

To explore variations in the data for identification, Table A.10 provides a summary of the empirical repurchasing rates based on past experiences. Conditioning on the total number of experiences (controlling for household selection), the repurchasing probability, shown in the last column, is more responsive to the satisfactory rating of the prior experience under laser than under sticker. For example, in the laser markets, households with 1 non-satisfactory experience is 0.521 less likely to repurchase than households with 1 satisfactory experience. However, in the sticker markets, the difference in repurchasing probability is only 0.045. This pattern is indicative of faster belief updating under laser, which aligns with the reduced-form results in Table 6.

6.3 Results

Table 7 Column 1 presents the estimated parameters using simulated maximum likelihood (ML). Column 2 restricts the sample to households with more than 6 watermelon purchases during the entire season, allowing for more robust identification of the learning parameters. The estimates are qualitatively and quantitatively robust. Column 3 tests a static model by shutting down belief updating and regime switching, and setting $\tilde{\gamma}_{imjt} = \underline{\gamma}$, for all $j = 1, 2, 3$. The likelihood ratio test (between Columns 2 and 3) rejects the static model against the dynamic learning model.

Taking estimates in Column 2, the estimated prior probability $\hat{\lambda}^0$ is 0.24 for laser and 0.04 for sticker. The estimated recovery rate is 0.502. These point estimates are consistent with the reduced-form results, suggesting that prior beliefs are more optimistic under laser labels than under sticker labels. Belief updating is faster with the improved prior: following one satisfactory experience, the posterior beliefs ($\tilde{\lambda}$) increase to 0.4 under laser but only to 0.09 under sticker.

The negative estimated values of $\hat{\tau}_s$ and $\hat{\tau}_l$ indicate that consumers tend to downgrade the normal pile when sellers offer it alongside another pile labeled as premium, especially under the sticker labels. This finding aligns with the experimental results in Table 3, which show a significant negative impact on the sales of the normal pile for the sticker group. The estimated $\hat{\eta}$ and $\hat{\eta}_l$ rationalizes the amount premium purchases in the sample, relative to that of the other options. The positive value of $\hat{\eta}_l$ suggests the possibility of a "foot-in-the-door" effect that interacts with purchases and learning. In other words, the higher initial take-up of the premium pile induced by the laser labeling may further accelerate the learning process and reputation building for the sellers offering premium watermelons.

6.4 Evolution of Beliefs

In this section, I use the structural estimates to analyze how beliefs evolve over time and how this affects seller's reputation incentives. Figure 6 displays model-simulated market average beliefs about the seller (λ) over time under different scenarios.

First, the circle-marked line shows the market average beliefs for the sample of households in the sticker markets, using the estimated sticker prior beliefs from Column 2 of Table 7, as well as the empirical prices and quality provided by sellers in the sticker markets. The empirical satisfaction rate among households in the sticker markets is 0.36, which is qualitatively similar to the satisfaction rate for the undifferentiated pile (0.3). The square-marked line maintains the same sticker prior but replaces the pricing and quality provision with that observed in the laser markets.[32] Comparing these two scenarios highlights the challenge of building reputation under the sticker label: after three seasons (21 weeks), market average beliefs under the latter (with higher quality provision) improve only modestly relative to the former.

Next, I examine the evolution of beliefs under the laser label. The diamond-marked line replaces the prior beliefs with that under the laser label but keeps the same pricing and quality provision as that under the sticker label. Finally, the asterisk-marked replaces both the prior beliefs and sellers' pricing and quality provision with that under the laser label. Comparing to the prior two scenarios with sticker priors, we see that, holding supply-side behavior fixed, the laser label alone has a significant impact on beliefs. This difference, in turn, affects sellers' incentives to provide quality, further driving markets to different outcomes over time. These counterfactual exercises also demonstrate two quantitatively important roles of the laser signal: first, it boosts the level of initial prior (comparing the y-intercept of the diamond-marked line to that of the circle-marked line), which stimulates higher initial demand. Second, conditioning on quality provision, laser labeling induces faster belief updating (comparing the slope of the asterisk-marked line to that of the square-marked line), which speeds up reputation building. Combining these forces, sellers who provide high quality under laser enjoy a significantly higher reputation by the end of the simulated period, compared to low quality provision under sticker, and the three-season discounted consumer surplus is 21% higher.[33]

[32] The empirical satisfaction rate among households in the laser markets is 0.528.

[33] 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 is given by:

$$E(CS_{it}) = \frac{1}{\alpha_0}\left[\log\left(\sum_{j=1}^{J}\exp(V_{ijt}(\tilde{\gamma}_{ijt}))\right) + \sum_{j=1}^{J}\tilde{\pi}_j(V_{ijt}(\gamma_{jt}) - V_{ijt}(\tilde{\gamma}_{ijt}))\right], \qquad \text{where} \quad \tilde{\pi}_j = \frac{\exp(V_{ijt}(\tilde{\gamma}_{ijt}))}{\sum_{k=1}^{J}\exp(V_{ijt}(\tilde{\gamma}_{ijt}))}$$

7 Conclusion

This study theoretically and empirically examines the lack of quality provision in a developing country retail market setting. I find that information frictions can hinder quality provision, and sellers' reputation incentive crucially depends on the dynamics of consumer learning. Introducing a costly signaling technology helps enhance consumer learning and induces reputation building. That said, small individual sellers may not have the incentive to invest in such expensive technologies themselves. Though the exact learning process and cost of reputation building are different for different products and markets, the study highlights a number of broad takeaways and directions for future research.

First, a good reputation takes time to establish. As countries develop and demand for quality increases, reputation for high quality may eventually emerge. However, in developing countries that lack such reputable entities, prevailing market beliefs matter for firms' incentive to invest in quality. Interventions that enhance consumer beliefs and facilitate learning can help to restore sellers' reputation incentives and benefit both sellers and consumers.

Second, many industries in developing countries are characterized by fragmented markets with a large number of small firms. Such market fragmentation can discourage quality provision as small firms may not find it profitable to undertake costly signaling activities that require large upfront costs.

Third, while the market-based reputation mechanism offers an alternative solution to address the information problem, as opposed to direct quality control, it may not function effectively in countries with weak regulatory institutions. In the context of China, pessimistic beliefs under sticker labels are partly due to rampant past counterfeiting activities. The discussion highlights a potential interaction between the market-based reputation mechanism and government regulations in establishing trust among consumers and strengthening firms' incentives to invest in high quality.

Finally, the current study focuses on sellers in the downstream markets and abstracts away from the role of the supply chain. One could imagine that once quality can be priced in downstream, such incentive may trickle up and induce quality production among the upstream producers. In general, how quality incentives are distributed along the supply chain and how that may affect the organization of quality production along the chain remain an open area for future research.

Data Availability Statement

The data underlying this article are available in Zenodo, at https://dx.doi.org/10.5281/zenodo.13909671

The second term in the outer bracket takes into account the fact that purchasing decisions are made under the current beliefs $\tilde{\gamma}_{ijt}$ whereas the true underlying quality is γ_{jt}.

References

Allen, T. (2014), 'Information frictions in trade', Econometrica **82**(6), 2041–2083.

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

Bai, J., Gazze, L. and Wang, Y. (2022), 'Collective reputation in trade: Evidence from the chinese dairy industry', Review of Economics and Statistics **104**(6), 1121–1137.

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 Duflo, E. (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 Tadelis, S. (2008), Seller reputation, Now Publishers Inc.

Bardhan, P., Mookherjee, D. and Tsumagari, M. (2013), 'Middlemen margins and globalization', American Economic Journal: Microeconomics **5**(4), 81–119.

Björkman Nyqvist, M., Svensson, J. and Yanagizawa-Drott, D. (2022), 'Can good products drive out bad? a randomized intervention in the antimalarial medicine market in uganda', Journal of the European Economic Association **20**(3), 957–1000.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D. and Roberts, J. (2013), 'Does management matter? evidence from india*.', Quarterly Journal of Economics **128**(1).

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

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

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

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

Grace, D., Roesel, K. and Lore, T. (2014), 'Food safety in informal markets in developing countries: An overview'.

Green, E. J. and Porter, R. H. (1984), 'Noncooperative collusion under imperfect price information', Econometrica: Journal of the Econometric Society pp. 87–100.

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

Jensen, R. and Miller, N. H. (2018), 'Market integration, demand, and the growth of firms: Evidence from a natural experiment in india', American Economic Review **108**(12), 3583–3625.

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

Kugler, M. and Verhoogen, E. (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 Morjaria, A. (2015), 'The value of relationships: evidence from a supply shock to kenyan rose exports', The American Economic Review **105**(9), 2911–2945.

Michelson, H., Fairbairn, A., Ellison, B., Maertens, A. and Manyong, V. (2021), 'Misperceived quality: Fertilizer in tanzania', Journal of Development Economics **148**, 102579.

Pei, H. (2023), 'Reputation building under observational learning', The Review of Economic Studies **90**(3), 1441–1469.

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

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

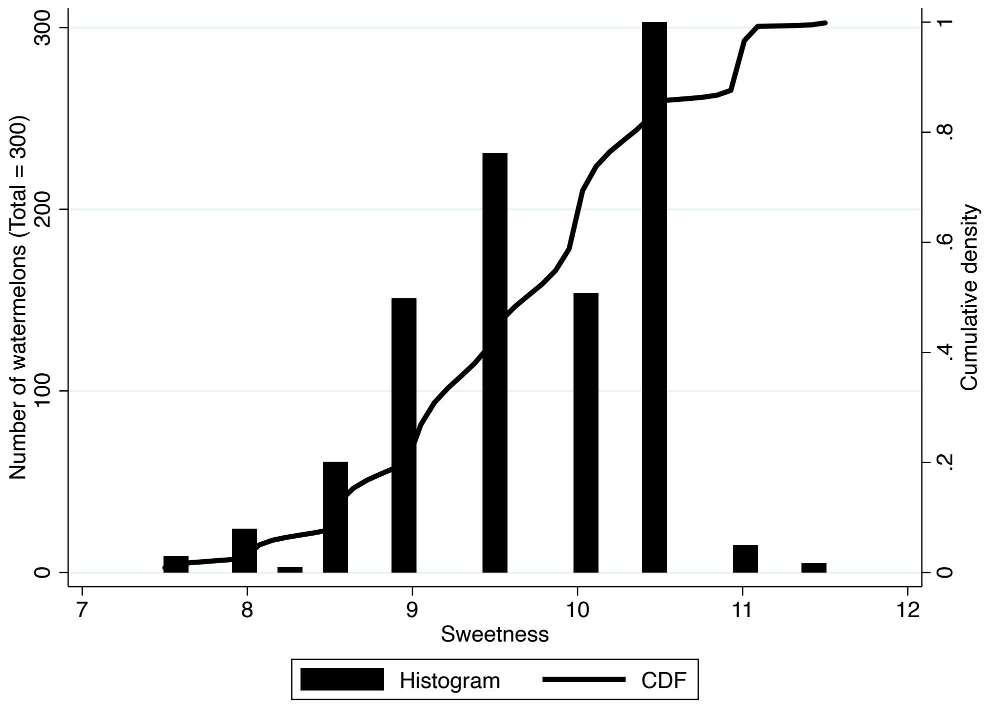
Train, K. E. (2009), Discrete choice methods with simulation, Cambridge University Press.

Verhoogen, E. (2021), 'Firm-level upgrading in developing countries'.

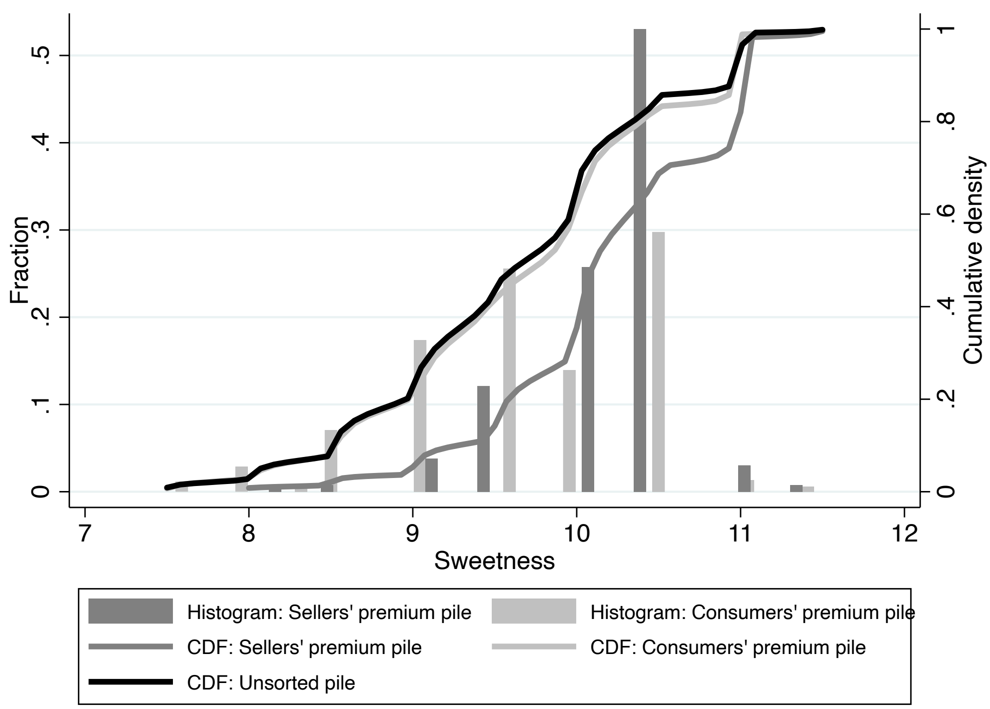
Verhoogen, E. A. (2008), 'Trade, quality upgrading, and wage inequality in the mexican manufacturing sector', The Quarterly Journal of Economics **123**(2), 489–530.

Figure 1: Variation in Quality and Asymmetric Information between Sellers and Consumers

Panel A.

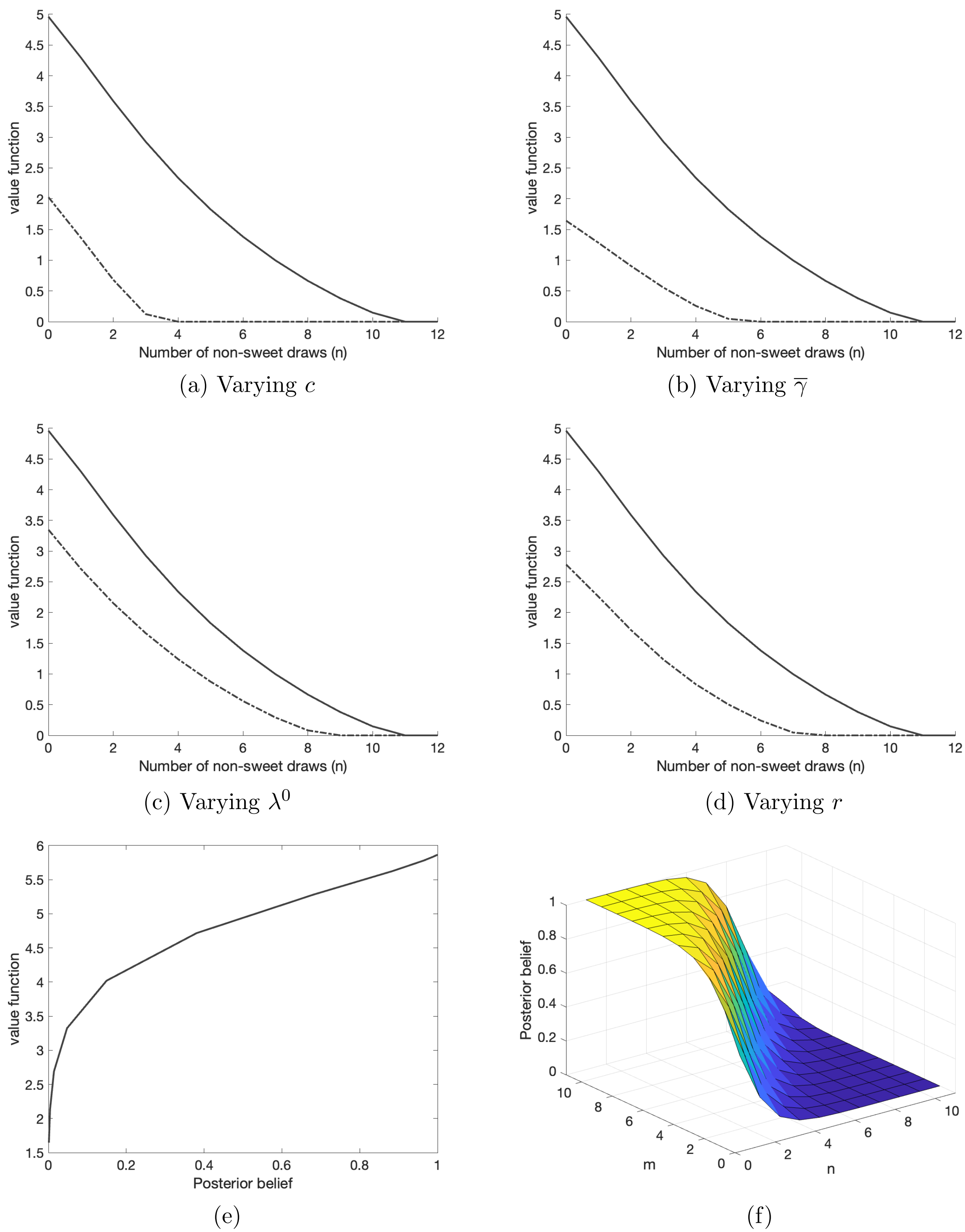


Panel B.



Note: This figure shows the empirical distributions for (1) all 300 randomly picked watermelons used in the sorting tests (Panel A) and (2) the premium piles sorted by sellers and consumers (Panel B). 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 B.1.

Figure 2: Theoretical Model Simulation: Return of Building Reputation



(a) Varying c (b) Varying $\overline{\gamma}$

(c) Varying λ^0 (d) Varying r

(e) (f)

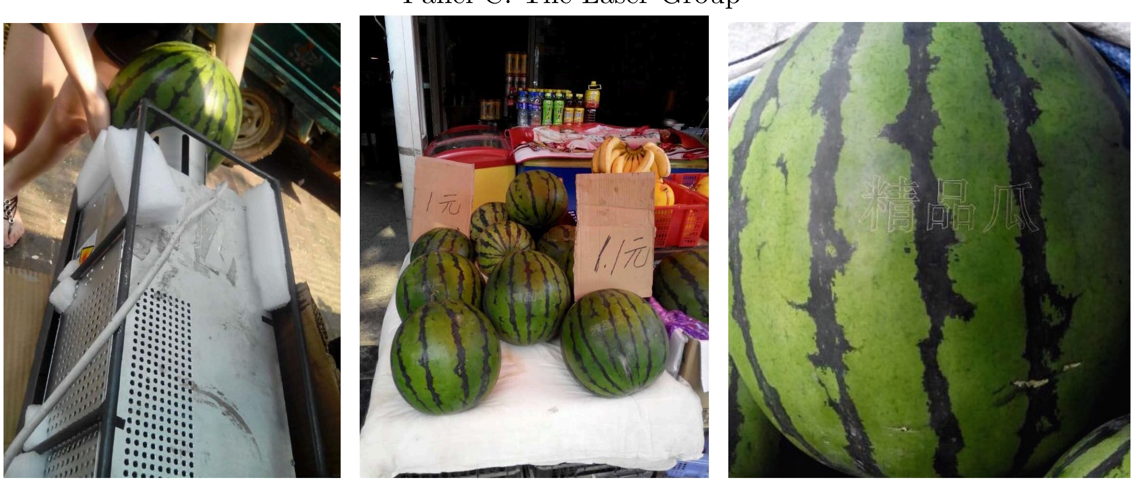
Note: This figure simulates the theoretical model in Section 3 under different parameter values. The solid line in (a)-(d) simulates the model under the following set of parameter values: $c = 0.1, \underline{\gamma} = 0.3, \overline{\gamma} = 0.8, \lambda_0 = 0.5, r = 0.5, \delta = 0.95$. The dotted line in (a) changes $c = 0.3$; (b) changes $\overline{\gamma} = 0.6, \underline{\gamma} = 0.3$; (c) changes $\lambda_0 = 0.05$; (d) changes $r = 0.1$. Every line in (a)-(d) plots V_G^+ against different values of n, holding m fixed at 0. Panel (e) plots V_G^+ against $\tilde{\lambda}$ and Panel (f) plots $\tilde{\lambda}$ against m and n, under the same parameter values as the solid lines in (a)-(d).

Figure 3: Pictures of the Labeling Treatments

Panel A. The Label-Free Group

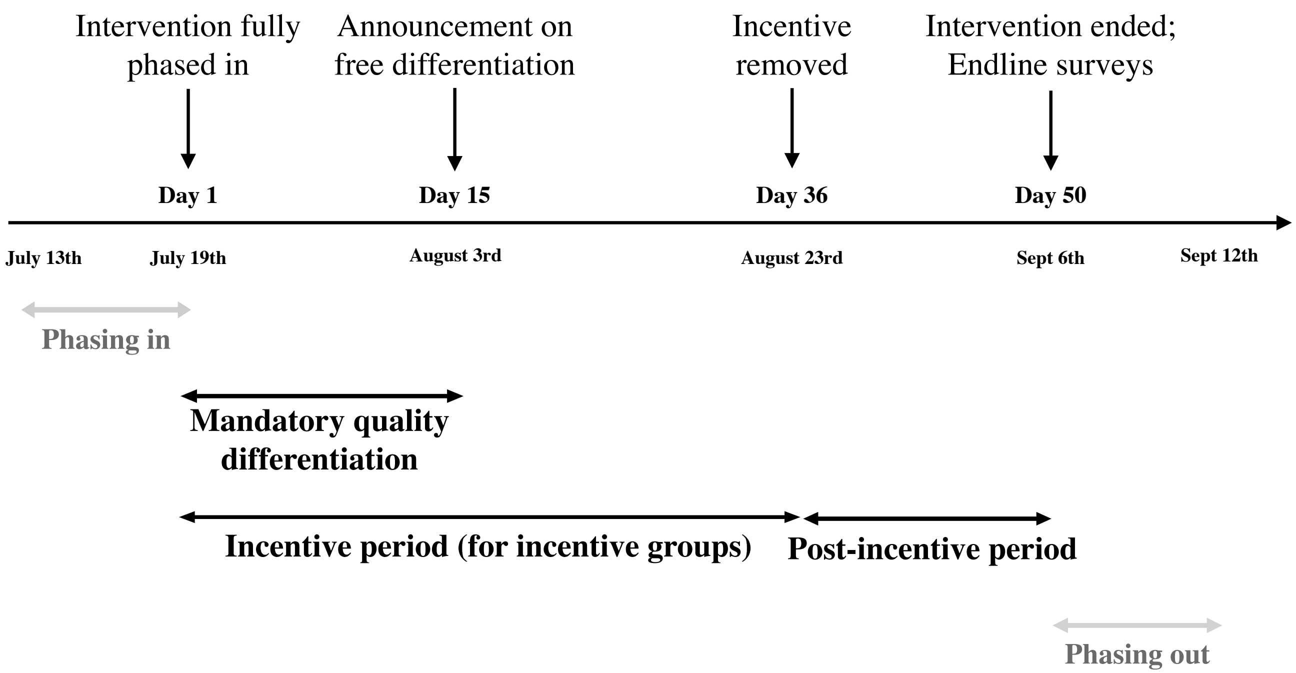
Panel B. The Sticker Group

Panel C. The Laser Group



Note: This figure depicts the actual implementation of the labeling 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 labeled the watermelons in the premium pile. Nothing was done for the label-free 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 4: Timeline of the Intervention

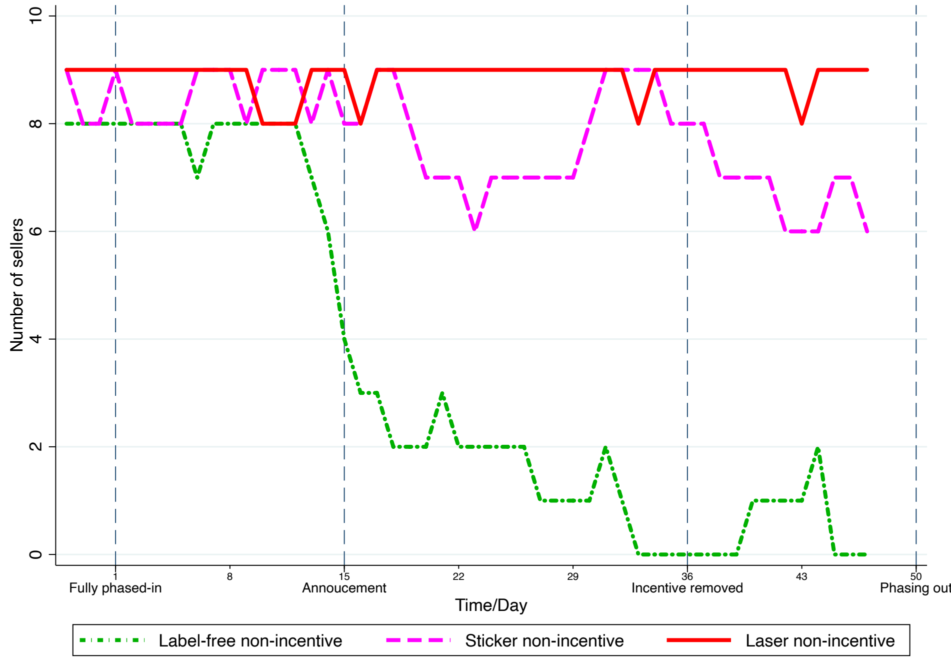


Note: This figure gives an overview of the timeline of the study.

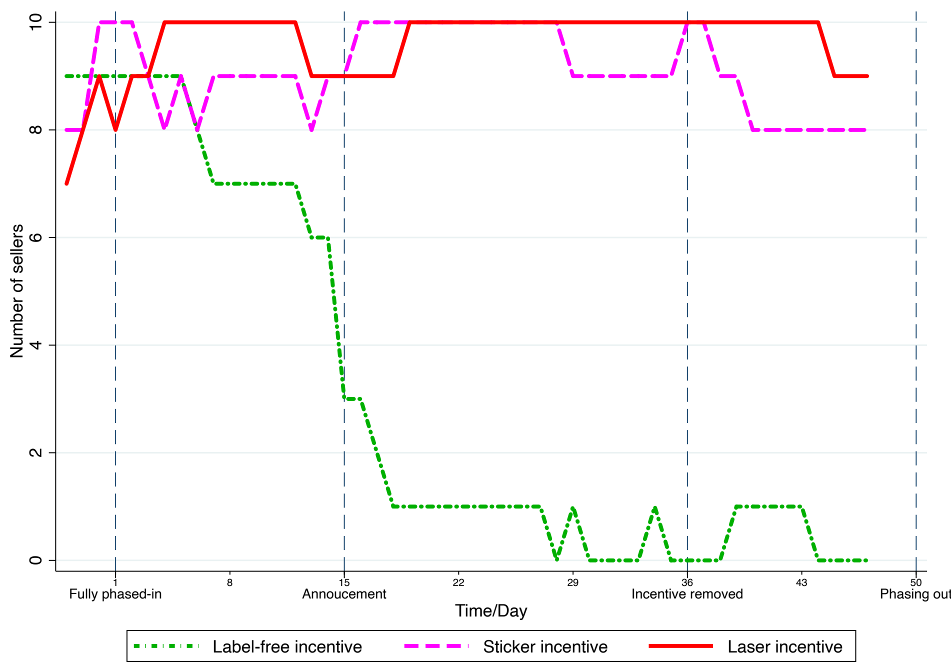
1. The intervention was rolled out from July 13 to 19, 2014.
2. All sellers were asked to experiment with quality differentiation for the first 2 weeks, from July 19 to August 3. To participate in the experiment, sellers signed an agreement form at the beginning of the period that they would experiment 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 thereafter.
3. On August 23, 35 days (6 weeks) into the intervention, the incentive (for the incentive groups) was lifted.
4. September 6 was the last day of the intervention. An endline survey was conducted at surveyors' final visits to sellers' stores. Most of the data analysis focuses on the period from July 19 (day 1) to September 6 (day 50).
5. The market intervention was gradually phased out from September 6 to September 12, 2014.
6. A follow-up survey was conducted from September 14 to 20, 2014, and another one was conducted a year later, in July 2015.

Figure 5: Quality Differentiation at Sale

Panel A. Non-Incentive Groups

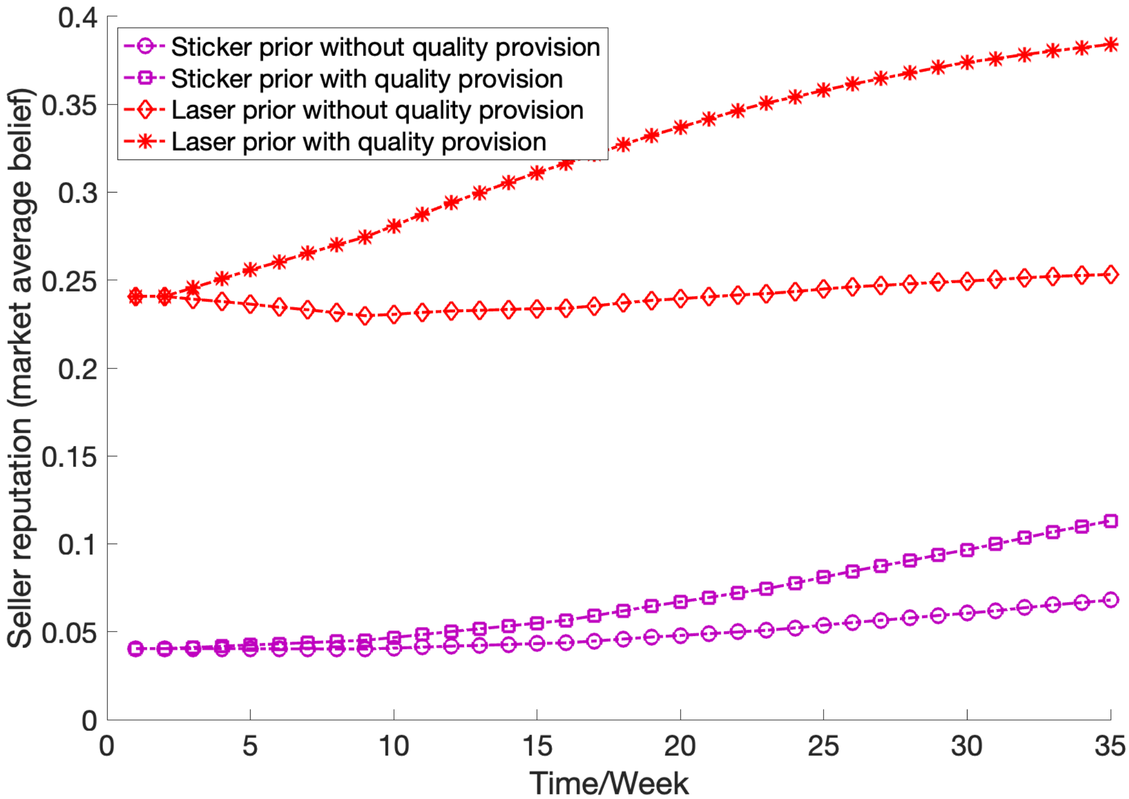


Panel B. Incentive Groups



Note: This figure plots the number of sellers who differentiated quality at sale in each treatment group over time. Panel A shows that for the non-incentive groups and Panel B shows that for the incentive groups. The time axis runs from July 19 (day 1) to September 6 (day 50), 2014, corresponding to the period of the fully phased-in 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-free group reneged from the beginning.

Figure 6: Evolution of Beliefs



Note: This figure illustrates the model-simulated evolution of market-average beliefs (λ) over time under various scenarios. The circle-marked line represents the market average beliefs for the sample of households in the sticker markets using the estimated sticker prior (from Table 7), along with the prices and quality provision observed in the sticker markets. The square-marked line maintains the same sticker prior but replaces the pricing and quality provision with that observed in the laser markets. The diamond-marked line replaces the prior beliefs with that under the laser label but keeps the pricing and quality provision the same as that for the sticker markets. Finally, the asterisk-marked line replaces both the prior beliefs and sellers' pricing and quality provision with that observed under the laser label.

Table 1: Baseline Summary Statistics

	Observations	Median	Mean	Std Dev
Panel A. Community and Market Characteristics				
Size measured in number of housing units	60	1350.000	1915.133	1930.216
Housing price (in thousand RMB/meter^2)	60	8.950	8.291	1.594
Fraction elderly residents	60	0.250	0.280	0.123
Distance to the nearest supermarket (in kilometers)	60	1.500	1.567	1.046
Years since establishment	60	15.500	17.633	11.242
Number of competitors in the local market	60	3.000	3.533	2.273
Panel B. Seller Characteristics				
Gender (female=1 and male=0)	60	0.000	0.483	0.504
Age	60	42.000	41.067	9.189
Years of schooling	59	9.000	10.254	2.509
Selling fruit as primary income source (dummy)	60	1.000	0.950	0.220
Selling fruit only in the summer (dummy)	60	0.000	0.033	0.181
Planning to stop selling fruit (dummy)	60	0.000	0.017	0.129
Number of years selling fruit	60	8.000	9.017	6.035
Number of years selling fruit at this location	60	6.500	7.867	6.239
Planning to relocate (dummy)	60	0.000	0.000	0.000
Purchasing from fixed wholesaler(s) (dummy)	60	0.000	0.217	0.415
Panel C. Household Characteristics				
Household size	658	3.500	3.760	1.366
Fraction of elderly residents	657	0.000	0.169	0.272
Fraction female residents	657	0.500	0.498	0.154
Household monthly income (in thousand RMB)	647	4.000	5.250	3.235
Fruit as % of total food consumption	602	30.000	32.010	17.906
Watermelon as % of total fruit consumption	626	30.000	35.627	25.292
Number of watermelons consumed per week	654	1.000	1.308	0.695
Local markets as main purchase source (dummy)	675	1.000	0.756	0.430
Supermarkets as main purchase source (dummy)	675	0.000	0.227	0.419
Willingness to pay for quality (RMB/Jin)	633	2.000	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 is due to missing responses in the baseline surveys. To elicit willingness to pay for quality, households were asked 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: Effects of Labeling Treatment on Quality Provision

Dep. var.: Quality measured in sweetness

Panel A. Quality of the Premium Pile

	All		Non-Incentive Only	
	(1)	(2)	(3)	(4)
Laser	0.509***	0.418**	0.711***	0.619**
	(0.176)	(0.176)	(0.222)	(0.266)
Observations	468	468	238	238
Baseline controls		Yes		Yes
Time fixed effects	Yes	Yes	Yes	Yes
Omitted group mean	10.184		9.738	
Std. dev.	1.102		1.104	
Small sample robustness				
Permutation test (p-value)	0.004	0.026	0.003	0.020
Clustered bootstrap (p-value)	0.004	0.027	0.001	0.085

Panel B. Quality Differentiation Behavior

	Sweetness level		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 effect	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Normal pile mean	9.787	9.366	0.102	-0.285
Std. dev.	0.990	0.923	0.774	0.965
Small sample robustness				
Clustered bootstrap (p-value)	0.000	0.009	0.000	0.002

Note: This table examines quality provision by treatment group. Quality is measured in sweetness. In Panel A, each observation is at the seller-biweekly (every quality sampling check) level. Columns 1 and 2 include all sticker and laser markets, and Columns 3 and 4 restrict the sample to the non-incentive groups. All regressions include time (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. In Panel B, each observation is at the seller-pile-biweekly level. The key explanatory variable is a dummy for the premium pile (the omitted group is the normal pile). The dependent variable for Columns 1 and 2 is the level of sweetness, and that for Columns 3 and 4 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-free group at each quality sampling visit (from week 3 onward). All regressions in Panel B include seller and time fixed effects. Standard errors are clustered at the seller level. The small sample robustness check implements two different procedures to address concerns over the relatively small sample size. In Panel A, a permutation test reports the p-values for the test of the null hypothesis that laser has no effect by randomly permuting the values for the laser dummy 1,000 times while respecting seller clusters. The clustered bootstrap method is used to perform nonparametric bootstrap estimation of the regression coefficients. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 3: Effects of the Labeling Treatments on Price, Sales and Profits

	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**	0.046***	49.852*	49.454*	0.001	-0.001	-40.374	-55.550**	9.478	-6.096
	(0.199)	(0.196)	(0.016)	(0.016)	(28.758)	(28.506)	(0.010)	(0.010)	(24.860)	(23.831)	(39.378)	(41.676)
Laser	0.297*	0.396**	0.070***	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	1427	1427	1462	1462	1427	1427	1462	1462	1462	1462
Baseline controls		Yes		Yes		Yes		Yes		Yes		Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Small sample robustness												
Permutation test (p-value)												
Sticker	0.883	0.864	0.096	0.057	0.092	0.135	0.899	0.955	0.165	0.045	0.817	0.884
Laser	0.112	0.058	0.001	0.005	0.040	0.030	0.553	0.909	0.663	0.878	0.213	0.117
Clustered bootstrap (p-value)												
Sticker	0.876	0.860	0.019	0.015	0.080	0.120	0.898	0.955	0.113	0.035	0.804	0.894
Laser	0.061	0.026	0.000	0.003	0.006	0.010	0.541	0.901	0.659	0.835	0.188	0.078
Laser-free mean	4.254		0.044		47.104		-0.002		177.099		224.202	
Std. dev.	0.623		0.081		102.476		0.059		108.714		120.402	

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-free group, the mean and standard deviation for which are shown in the last two rows. Price Δ is defined as 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 differentiating 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 include day fixed effects. Standard errors are clustered at the seller level. Small sample robustness implements two different procedures to address concerns over a relatively small sample size. A permutation test reports the p-values for the test of the null hypothesis that laser (sticker) has no effect by randomly permuting the values of labeling treatment group assignment 1,000 times while respecting seller clusters. The clustered bootstrap method is used to perform nonparametric bootstrap estimation of the regression coefficients. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 4: Effects of the Incentive Treatment on Quality Provision

Dep. var.: Sweetness of the premium pile

	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		Yes		Yes

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., weeks 7 and 8). The key explanatory variable is the interaction term. Each observation is at the seller-biweekly (corresponding to each quality sampling visit) 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, 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 levels, respectively.

Table 5: Heterogeneity in Price and Quality by Sellers' Sorting Ability

	Premium price		Premium sweetness	
	(1)	(2)	(3)	(4)
	All	Laser	All	Laser
Panel A. Ability (dummy) measured at the sorting test				
ability dummy	0.068**	0.086*	0.201	0.410*
	(0.025)	(0.044)	(0.159)	(0.221)
log housing price	0.024	0.326***	-0.077	0.570
	(0.028)	(0.086)	(0.146)	(0.883)
log number of housing units	0.023	-0.005	-0.292**	-0.014
	(0.016)	(0.015)	(0.111)	(0.138)
Observations	1454	484	352	118
Group FE	Yes		Yes	
Time FE	Yes	Yes	Yes	Yes
Panel B. Ability (discrete) measured at the sorting test				
ability	0.029*	0.050	0.008	0.405***
	(0.014)	(0.036)	(0.078)	(0.092)
log housing price	0.024	0.392**	-0.096	1.558**
	(0.035)	(0.144)	(0.159)	(0.613)
log number of housing units	0.027	0.004	-0.308**	0.111
	(0.017)	(0.024)	(0.125)	(0.118)
Observations	1454	484	352	118
Group FE	Yes		Yes	
Time FE	Yes	Yes	Yes	Yes

Note: This table examines the heterogeneity in price and quality provision by sellers' sorting ability. Observation for price is at seller-day level and observation for quality is at seller-biweekly (corresponding to each quality sampling visit) level. Ability is measured based on sellers' performance in the sorting test. Panel A uses an ability dummy that equals to 1 if and only if the seller did not make any *clear mistake* in the sorting test. Clear mistake is defined to be a case in which at least one watermelon sorted to the low pile strictly dominates the quality of one (or more) sorted to the high pile. Panel B further separates those sellers who did not make clear mistakes into two categories: (1) those whose high pile weakly dominates the low pile (ability = 1): that is, the highest quality of the low pile equals to the lowest quality in the high pile; (2) those whose high pile strictly dominates the low pile (ability = 2). Columns (1) and (3) include all sellers and control for group fixed effect. Column (2) and (4) restrict to sellers in the laser group. All regressions control for time fixed effect. Standard errors are clustered at the seller level.

Table 6: Household Purchasing Dynamics under Different Labeling Technologies

	Households in the Laser Markets		Households in the Sticker Markets	
	(1)	(2)	(3)	(4)
Panel A. Decision to purchase from the premium pile				
Lagged avg. satisfaction rating	0.261***		0.049	
	(0.073)		(0.068)	
Lagged % of very good experiences		0.393***		0.110
		(0.090)		(0.123)
Observations	191	193	183	183
Household Baseline Controls	Yes	Yes	Yes	Yes
Week Fixed Effects	Yes	Yes	Yes	Yes
Panel B. Decision to purchase from of the normal pile				
Lagged avg. satisfaction rating	0.035		-0.014	
	(0.042)		(0.029)	
Lagged % of very good experiences		0.010		-0.016
		(0.076)		(0.051)
Observations	520	576	497	530
Household Baseline Controls	Yes	Yes	Yes	Yes
Week Fixed Effects	Yes	Yes	Yes	Yes

Note: This table examines the purchasing dynamics under different labeling technologies. Each observation is at the household-week level. The dependent variable for Panel A is whether the household has purchased any watermelons 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 or normal watermelons purchased prior to the period and (2) the percentage of past consumption experiences that attained the highest satisfaction rating of 5. All regressions include week fixed effects and control for the following set of household baseline characteristics: household size, percentage of elderly residents, 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 household level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 7: Simulated Maximum Likelihood Estimation Results

Parameters	Full Sample (1)	Frequent Sub-sample (2)	Static Model (3)
λ_s^0	0.124	0.040	
	(0.007)	(0.003)	
λ_l^0	0.279	0.241	
	(0.005)	(0.002)	
r	0.624	0.502	
	(0.017)	(0.024)	
α	0.173	0.136	0.192
	(0.106)	(0.000)	(0.001)
η	-2.674	-2.892	-2.697
	(0.088)	(0.003)	(0.007)
η_l	0.956	1.226	1.018
	(0.063)	(0.002)	(0.005)
τ_s	-1.907	-2.073	-1.961
	(0.134)	(0.003)	(0.004)
τ_l	-0.389	-0.186	-0.396
	(0.210)	(0.003)	(0.002)
Market FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Log-likelihood	-3051.4	-1123.0	-1199.6

Note: This table shows the simulated maximum likelihood estimation results. Column 1 shows the estimates using the full household sample. Columns 2 and 3 restrict to households with more than 6 purchases during the season. Column 3 estimates a static model without learning. Estimates for the market and time fixed effects are abbreviated. Standard errors shown in parentheses are calculated as the square root of the inverse of the Hessian matrix.