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

Jie Bai (Harvard Kennedy School) †

July 2023

CLICK HERE FOR LATEST VERSION

Abstract

Quality provision is often low in many developing-country markets, and firms commonly lack a reputation for quality. This paper examines this problem both theoretically and empirically, in the context of the retail watermelon markets in China. I first demonstrate the existence of substantial asymmetric information on quality between sellers and buyers, as well as a stark absence of a quality premium at baseline. To explain this, I develop a theoretical model that highlights the influence of consumer beliefs and costly signaling on shaping sellers' reputation incentives. I then conduct an experiment by randomly introducing two signaling technologies into different markets: a cheap sticker label and an expensive laser-cut label. The findings reveal that the laser label prompts sellers to offer higher quality, resulting in increased sales profits, while the sticker label fails to achieve the same effect. Leveraging the experimental variation, I estimate an empirical model of consumer learning to uncover the underlying evolution of beliefs. The results demonstrate that pessimistic beliefs under the sticker label can impede reputation building. On the other hand, 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 problem 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 such market environments, characterized by information frictions and mistrusts among consumers, firms require a good reputation to thrive. However, a reputation for quality is precisely what many firms in developing countries lack. 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 becomes: what impairs firms' incentive and ability to establish a reputation for quality? Understanding this can inform solutions to foster reputation building and improve quality provision.

I theoretically and empirically examine this problem in the context of the retail watermelon markets in China. I first demonstrate substantial asymmetric information on quality between sellers and buyers and 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' reputation incentive. The theoretical analysis shows that pessimistic beliefs can make reputation building a low-return investment, and markets can become stuck in a bad equilibrium despite a high demand for quality. Introducing costly signals can help enhance consumer learning and restore sellers' reputation incentive. Motivated by the theoretical discussion, I then set up an experiment aimed at inducing quality provision and reputation building by introducing different signals to the markets: a cheap sticker label and an expensive laser-cut label. Consistent with the theoretical model, the laser label serves as a costly signal and induces sellers to provide higher quality and earn higher sales profits, whereas the sticker label does not. Finally, leveraging the experimental variation, I estimate an empirical model of consumer learning to recover the underlying evolution of market beliefs. The results show that pessimistic beliefs can hinder reputation building and result in a sizable loss of consumer surplus.

The study takes place in local retail markets within a major Chinese city. These markets represent a typical setting found in 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 is the case with many food products and experience goods, assessing the quality of a watermelon at the point of purchase is challenging (short of cutting it open and tasting it). Fortunately, post-purchase, sweetness can serve as an indicator of quality and can be objectively measured using a sweetness meter. Leveraging this feature, I begin

[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).

by documenting substantial variation in watermelon quality at the baseline and a clear absence of a quality premium. In this market, sellers offer undifferentiated piles of watermelons at identical prices, with no seller consistently known for offering sweet watermelons at a higher price than others. To fully set up the puzzle, I conduct a sorting ability test with both sellers and buyers. The results show that sellers do possess the ability, though not perfect, to assess the quality of their watermelons, far better than consumers. This leads to the central question in the paper: given the information and ability to sort, and the repeated interactions between sellers and buyers in the market, why do sellers not capitalize on their ability to offer a quality premium and build a reputation for quality? Why is it that we do not see a quality premium or a reputation for quality emerging in these markets?

To examine why the reputation mechanism fails and what it takes to make it work, I first develop a theoretical model that highlights the importance of consumer beliefs and learning in shaping sellers' reputation incentive.[2] The model features a repeated game between a long-run seller and an infinite sequence of potential buyers. Sellers are of two types, low cost and high cost, and the former can spend efforts to improve the quality of the watermelons. However, due to the information problem, a seller's type and their claim of offering high quality cannot be immediately verified and need to be learned over time. To model the learning process, I assume that consumers hold a common prior belief about the seller's type and update their beliefs based on past consumption experiences in a Bayesian manner. To sustain the incentive to exert effort (as types are revealed), the model also features switching between good and bad regimes, similar to the punishment in Green and Porter (1984).

The theoretical analysis focuses on the reputation-building incentive of the low-cost type and shows that reputation building is history dependent: bad histories and pessimistic prior beliefs can make reputation-building a low-return investment. As a result, markets can become stuck in a low-quality-low-reputation equilibrium despite a high demand for quality. On the other hand, good histories and optimistic beliefs enhance the return of building reputation. I then use the model to investigate the role of costly signaling. The equilibrium analysis establishes a bound on the cost of the signal in order to achieve separation of types: the signal needs to be costly enough to deter the high-cost type from mimicking the low-cost type. There also exist a range of partial pooling equilibria, in which a positive fraction of the high-cost type adopts the signal (mimicking the low-cost type), but the fraction decreases as the cost of the signal increases.

The theoretical analysis highlights the role of costly signals in potentially influencing consumers' beliefs, where more optimistic beliefs are associated with costlier signals. This, in turn, may strengthen sellers' reputation incentive and lead to quality provision. Motivated by the theoretical insights, I designed an experiment that randomly introduces different signaling technologies into various markets. The experiment involved 60 sellers operating in 60 different local markets in Shijiazhuang, China. I

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

randomly introduced one of two signaling technologies into 40 out of the 60 markets: a cheap sticker label and a new expensive laser-cut label. Pilot surveys suggested that consumers perceive laser-labeling as an expensive and credible signal of quality. For a cross-randomized subset of sellers, I further provided a temporary monetary incentive to invest in high quality. This incentive treatment aimed to subsidize sellers' initial reputation building and shed light on the history-dependence aspect of reputation building. The intervention spanned eight weeks, encompassing the entire peak season for watermelons. To gather data from the supply side, I collected high-frequency quality, price, and sales data through daily field surveys. Additionally, biweekly quality sampling was conducted, with enumerators randomly sampling watermelons from the sellers and performing independent quality testing using sweetness meters. From the demand side, I recruited 675 households from the local markets and collected detailed household fruit purchase and consumption diaries over the entire intervention period to examine the demand side's responses.

The experiment yielded three main findings. First, laser labeling induced sellers to genuinely offer a quality premium, confirming the existence of reputation incentives that can potentially drive quality provision. Conversely, the sticker group did not demonstrate significantly higher quality or prices than the market average. Second, the incentive treatment successfully motivated sellers to provide higher quality, but this improvement was sustained only for the laser incentive group after the incentive was removed. Third, sellers in the laser group earned 30-40% higher sales profits on average, attributed to both higher prices and increased sales. However, the sticker group did not outperform the baseline. All the together, the results demonstrate a clear demand for quality, revealing that reputation can indeed be a rewarding investment. That said, in the subsequent season, when the free laser labeling was no longer provided, all markets reverted to the baseline situation, lacking quality differentiation. This suggests that small individual sellers may not have sufficient incentive to invest in the expensive technology themselves. The results also indicate a profitable entry opportunity for a large upstream firm capable of investing in the laser technology and gradually building up a reputation over time.

The experimental findings offer an explanation for the lack of quality provision observed at baseline. Additional evidence exploring the dynamics of household purchasing patterns and sales trajectories further highlights the role of consumer beliefs and learning in shaping sellers' reputation incentives. I next 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 setup of the theoretical model. To estimate the model, I employ simulated maximum likelihood, exploiting individual household purchasing decisions and reported consumption experiences for identification. Consistent with the experimental findings, the structural estimates indicate that consumers' prior beliefs are more pessimistic under the sticker than under the laser. Consequently, reputation can take a long time to establish, which explains why sellers do not have the incentive to provide quality at baseline. In contrast, the laser label enhances prior beliefs and learning, thereby strengthening sellers' reputation incentive. 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 with pessimistic beliefs and 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 relevant to developing countries that lack reputable entities and are dominated by small-scale firms. In such an environment, 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 the 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.[3] While many studies examine online trading environments, empirical work in the offline world is 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 discussed in Bar-Isaac and Tadelis (2008), the empirical challenge is that researchers typically do not observe all information available to buyers or sellers' behavior beyond what buyers observe. This study takes advantage of a field experiment that keeps track of both sides of the markets. The results demonstrate that consumers' beliefs and the way they gather information and learn shapes sellers' reputation incentive. This echoes the finding in Björkman Nyqvist et al. (2022) that quality provision of anti-malaria drugs in Uganda is hampered by consumer misconceptions. Although the contexts differ, the key takeaways are alike. To motivate high quality provision, mechanisms that enhance consumer learning or 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,[4] 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, which is the information problem and mistrust. Such mistrust, often targeted at a broad group level (for example, country-industry), generates an important externality that can hinder individual firms' incentive to upgrade quality and build reputation (Macchiavello, 2010; Bai et al., 2022).

[3]Prior studies have shown that information frictions matter in a variety of market settings (e.g., see).

[4]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).

The remainder of this paper is organized as follows. Section 2 describes the setting. Section 3 outlines the model. Section 4 describes the experimental design and the data. Section 5 presents the experimental results. Section 6 estimates an empirical model of consumer learning to recover the underlying evolution of market beliefs (seller reputation). Section 7 concludes the paper.

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).[5] 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.[6] 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.[7] 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[8]; the other, premium-quality pile sells at a higher price.[9] Figure A.3 plots the empirical distributions

[5] 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.

[6] 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.

[7] 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.

[8] 1 jin $\approx$ 1.1 pounds. The rest of the paper uses jin as the pricing unit.

[9] 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.

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 variation, 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.[10] This leads to the central question in the paper: considering all the conditions that appear conducive to

[10] 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.

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 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".[11]

To shed light on this puzzle, I first 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. Let's imagine 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 quality-improving effort, 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 with an additional 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. 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 to sort is 0. For a low-cost seller who is known to exert effort to sort, the resulting expected value of her watermelon is $\overline{\gamma}$. Assuming buyers pay their expected valuation, the resulting current-period profit of the seller is given by $(\overline{\gamma} - \underline{\gamma}) - c > 0$.

Play is assumed to proceed in two regimes: Good and Bad. In the Good regime, buyers expect a low-cost seller to exert effort, while in the Bad regime, the low-cost seller is expected 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

[11] 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."

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. This is needed to sustain effort in the long run, similar to the punishment along the equilibrium path in Green and Porter (1984).

Finally, to complete the baseline setup, suppose the play 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. Assume that each buyer observes the entire past sequence of draws and updates her belief about the seller's type in a Bayesian manner. Given the setup, 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, the 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 one is expected to exert effort. Market price in any given period is determined by the expected value.

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. Note that given the belief updating process described above, the seller would not have the 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.

Before presenting the full equilibrium analysis in Section 3.3, I first derive the seller's value functions and examine some properties. Let $V_G(m, n)$ denote the expected value to 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 expected value starting from a period in the Bad regime. We can write:

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

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

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:

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

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 belief of the low-cost type 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 equilibrium 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

The above discussion highlights the role of market beliefs in affecting the seller's reputation-building incentive (i.e., the incentive of the low-cost type to exert effort). Now I examine the equilibrium implications. The analysis focuses on the role of costly signals and examine under what conditions such a signal can induce effort to sort for the low-cost type and achieve separation between types.

I focus on 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.

[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.

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.

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:

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

$$V_B = \delta r V_G + \delta(1 - r)V_B \tag{6}$$

Re-arranging, we get:

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

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 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 $r^* = \frac{\delta(\overline{\gamma}-\underline{\gamma}+\underline{\gamma}c)-c}{\delta c}$.

[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.

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 Section 2. The model 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

[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):

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

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.

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 model 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 discussions, 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 theoretical analysis in Section 3.

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 expensive 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,

[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).

[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.

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 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.

[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.

[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.

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 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

[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.

[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.

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 and peaches (the second most popular fruit in the summer) by quality category.[24] Throughout the intervention, a total of 60,806 transaction records were collected, with 81% of transactions being for watermelons and 19% for peaches. 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 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

[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.

total, there are 15,292 purchase records, with 30.8% being 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.

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 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 during the experiment may not exactly align with 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 genuine 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 tech-

[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.

nologies 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 consider 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 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

[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.

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) 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

[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.

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 suggest that prior beliefs under stickers are likely to be pessimistic, consistent with discussion in Section 4.1. In such an environment, laser labeling improves both the prior mean and variance at the same time, both of which enhance the speed of reputation building.

Panel B repeats the same analysis for purchases from the normal pile. Since consumers are accustomed to purchasing unlabeled watermelons, each additional experience is not expected to shift their beliefs and significantly influence future purchase decisions. 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-building incentives. Why, then is there a lack of quality differentiation at baseline? In other words, did sellers have the incentive to adopt the costly signaling technologies themselves, and if so, why did they not do 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 one summer season), meant 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. Consequently, sellers were discouraged from adopting the costly laser technology or similar technologies at baseline. In general, a signaling technology needs to incur substantial upfront costs to be credible. However, these high fixed costs can present a significant barrier to adoption among small-scale sellers in developing countries. Instead, larger upstream wholesalers may be better positioned to invest in such technologies and build a brand for premium products.[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

[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.

[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.

buy the labels, rendering the signal worthless. This issue is similar for 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 function similarly 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 related 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 tracked-down sellers continued with quality differentiation. Additionally, examining data on peach sales did not show any significant impact on laser sellers' non-watermelon sales (as shown in Table A.9). This finding 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, signifying a wholesaler supplying premium watermelons under the laser brand. This belief is plausible since all sellers in the local markets source watermelons from the same wholesale market in the city. Moreover, the laser machine's cost is prohibitive for small local retailers, as discussed earlier, further supporting the perception of an upstream branding effect. In this context, consumers would be learning about the laser-labeled watermelon "brand" rather than the specific local retailer. Once sellers no longer carry that brand, consumers may no longer believe that sellers provide higher quality, and consequently, there would be no incentive for sellers to supply higher quality watermelons.

These findings highlight the challenges of reputation-building in supply chains where goods change hands multiple times (from farmers to traders to wholesalers and to local retailers) and every party along the supply chain may alter quality. The theoretical framework presented earlier focused on a single seller's reputation, and additional complexities arise in multi-layered supply chains. Future research is needed to explore these dynamics further and understand how reputation incentives can be effectively established and maintained in such settings.

5.7 Alternative Explanations

One alternative explanation for the success of laser labeling is that it is perceived as "cool" or visually appealing, which might add some direct utility of consumption. However, this "coolness effect" alone would not fully explain the observed household purchasing dynamics discussed in Section 5.5, as it goes beyond a static "coolness effect" and support a learning story.

To examine this further, I exploit the cross-randomized incentive treatment and compare sellers' sales dynamics within the same labeling treatment group. As we observed earlier in Table 4, the incentive treatment led to higher quality provision. Now, in Figure A.8, we can see that sales performance in fact diverges over time between the laser incentive and laser non-incentive groups. This divergence aligns with the earlier finding that the former provided higher quality, and over time, higher quality products yield better sales as consumers experience the product and update their perceptions. To quantify the differential dynamics more formally, Table A.10 estimates a linear time model and finds significant positive coefficients for the interaction terms between the incentive treatment and time for the laser group. Interestingly, we do not see such a dynamic pattern within the sticker groups, which is consistent with the household purchasing patterns analyzed in Section 5.5. The result again suggests that consumer beliefs update more slowly under the sticker technology, and as a result, the incentive treatment did not have a significant impact on the sales dynamics for sellers in the sticker group.

It is also important to acknowledge the role of relationships, which commonly exist in these markets (Fafchamps, 2002), and how that may affect the interpretation of the findings. For instance, sellers may selectively offer higher quality watermelons to repeat customers. Under this scenario, the lack of explicit quality differentiation at baseline would not be problematic if relational contracting perfectly allocated high-quality watermelons to high-valuation customers. However, if this 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 demand model of 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. Buyers make their 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. Specifically, I calculate the fraction of watermelons with sweetness above 10.5 (the same threshold used for the incentive). This gives a value of $\underline{\gamma} = 0.3$.[31] 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 assume $\overline{\gamma} = 0.625$ in the baseline estimation, and perform robustness checks with alternative values of $\overline{\gamma}$.

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.[32] 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, and the

[31] Figure A.9 plots the empirical distribution of sweetness of watermelons sampled from the label-free group.

[32] 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.

price coefficient α following standard arguments in the discrete choice literature. The key parameters of interest, λ_s^0, λ_l^0, and the recovery rate r, are identified from the dynamic household repurchasing decisions conditioned on their past experiences. Higher values of λ^0 not only lead to higher initial purchases but also potentially faster learning speed (for $\lambda^0 < 0.5$) and a larger increase in repurchasing probability following initial positive experiences. 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.11 provides a summary of the empirical repurchasing rates based on past experiences. For households in the laser group, the probability of repurchasing increases by approximately 63% when going from zero experience to one satisfactory experience. This increase is more pronounced compared to the sticker group, indicating faster belief updating, which aligns with the reduced-form results in Table 6. The repurchase probability following bad experiences informs the recovery rate r.

6.3 Results and Robustness Checks

Table 7 presents the estimated parameters using simulated maximum likelihood (ML). In Column 1, the estimation is based on the entire sample of 573 households. Column 2 restricts the sample to households with more than 5 watermelon purchases during the entire season (corresponding to the 25th percentile), 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 Column 2 as the preferred baseline estimates, the estimated prior probability $\hat{\lambda}^0$ is 0.24 for laser and 0.045 for sticker. The estimated recovery rate is 0.501. 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 new premium pile induced by the laser labeling may further accelerate the learning process and reputation building for the sellers offering premium watermelons. up learning and reputation building.

Table A.12 performs a number of additional robustness checks, including using alternative values of $\overline{\gamma}$ and excluding market fixed effects (to examine potential price endogeneity concerns). The results are quantitatively robust and similar to the baseline estimation in Table 7.

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 solid black (thin) 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 black (thin) dash-dotted line maintains the same sticker prior but replaces the pricing and quality provision with that observed in the laser markets. The empirical satisfaction rate among households in the laser markets is 0.528.Comparing these two scenarios highlights the challenge of building reputation under the sticker label: after three seasons (21 weeks), market average beliefs under the second scenario (with higher quality provision) do not improve significantly compared to the first.

Next, I examine the evolution of beliefs under the laser label. The solid red (thick) 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 red (thick) line with circles replaces both the prior beliefs and sellers' pricing and quality provision with that under the laser label. Comparing to the solid black (thin) line, we can observe that, even 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. In the end, sellers enjoy a significantly higher reputation under the last scenario compared to the first, and the three-season discounted consumer surplus is 21% higher.[33] These results illustrate the importance of costly signals in shaping consumer beliefs and the subsequent impact on seller reputation-building incentives.

[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:

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

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}.

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 instilling trust among consumers and strengthening firm reputation 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 passed through along the supply chain and how that affects the organization of quality production along the chain remain an open area for future research.

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'.

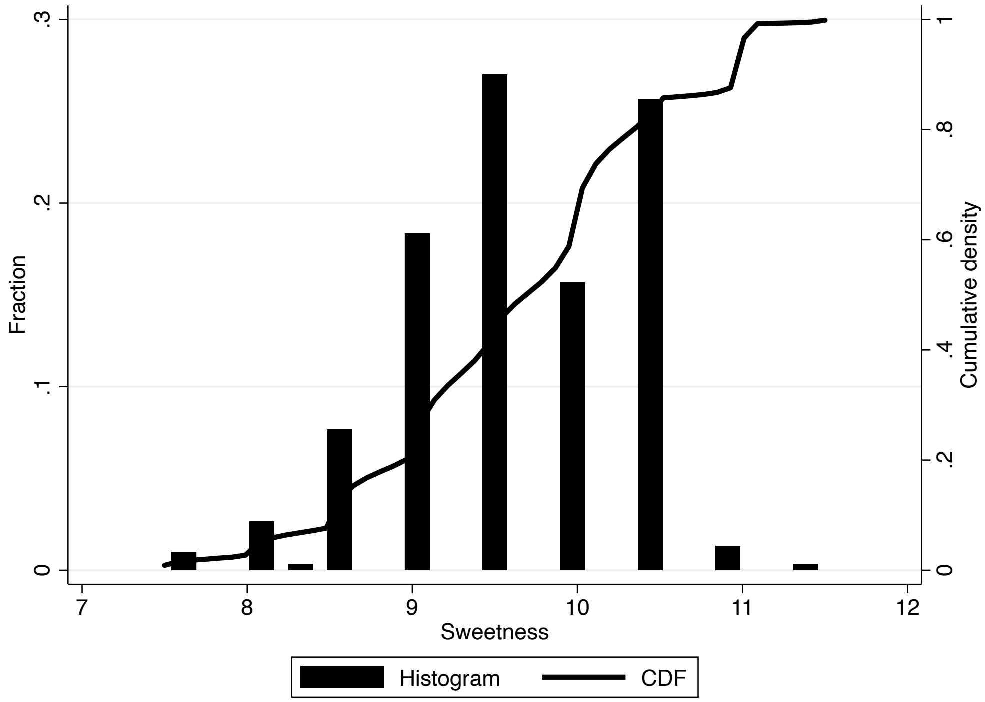
Train, K. E. (2009), Discrete choice methods with simulation, Cambridge University Press.

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

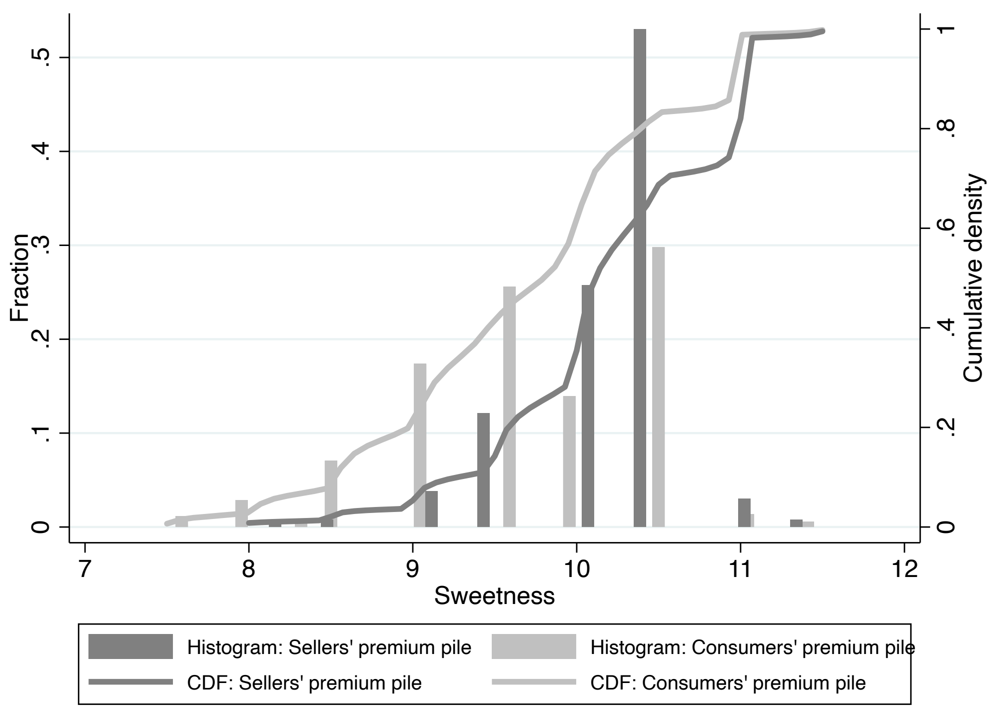
Verhoogen, E. A. (2008), 'Trade, quality upgrading, and wage inequality in the mexican manufacturing sector', The Quarterly Journal of Economics **123**(2), 489–530.

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

Panel A.

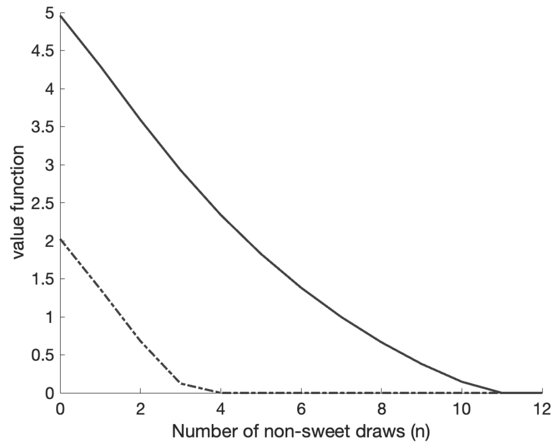


Panel B.

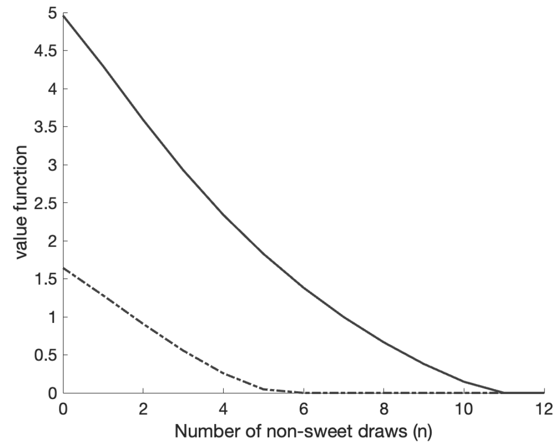


Note: This figure shows the empirical distributions for (1) all 300 randomly picked watermelons used in the sorting tests (Panel A) and (2) the premium piles sorted by sellers and consumers (Panel B). Quality is measured using a sweetness meter. For each watermelon, two measures are taken, one at the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

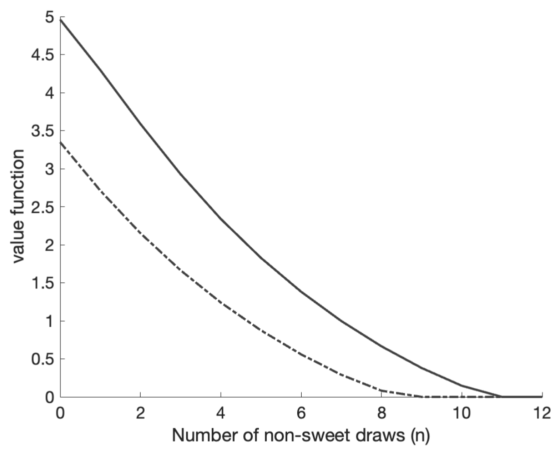
Figure 2: Theoretical Model Simulation: Return of Building Reputation



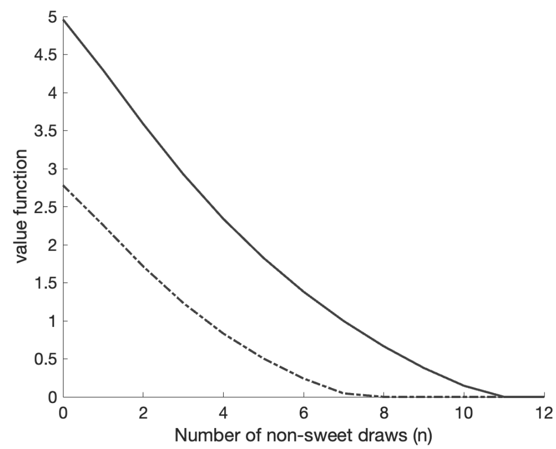
(a) Varying c



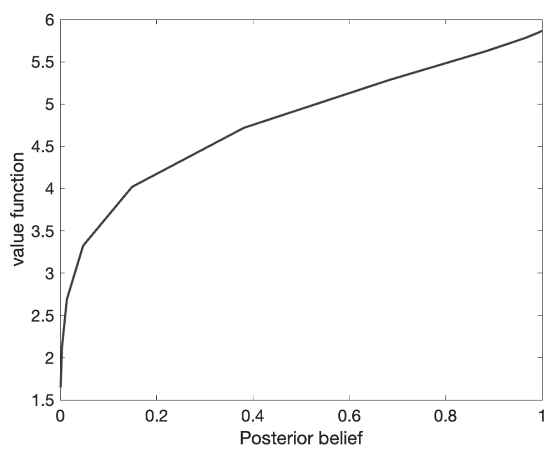
(b) Varying $\overline{\gamma}$



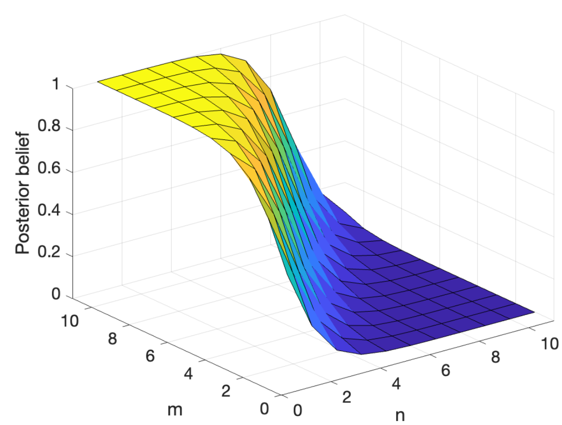
(c) Varying λ^0



(d) Varying r



(e)



(f)

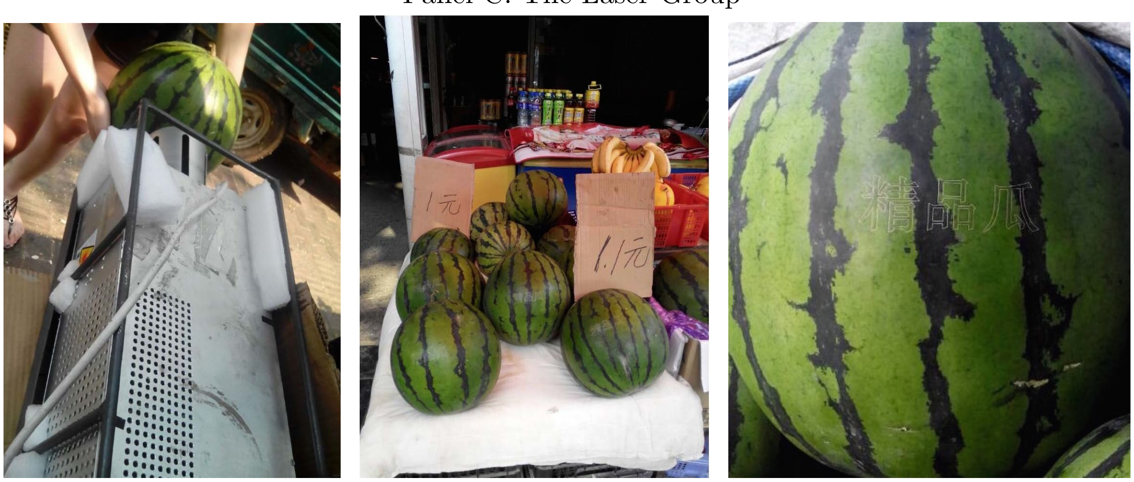
Note: This figure simulates the theoretical model in Section 3 under different parameter values. The solid line in (a)-(d) simulates the model under the following set of parameter values: $c = 0.1, \underline{\gamma} = 0.3, \overline{\gamma} = 0.8, \lambda_0 = 0.5, r = 0.5, \delta = 0.95$. The dotted 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

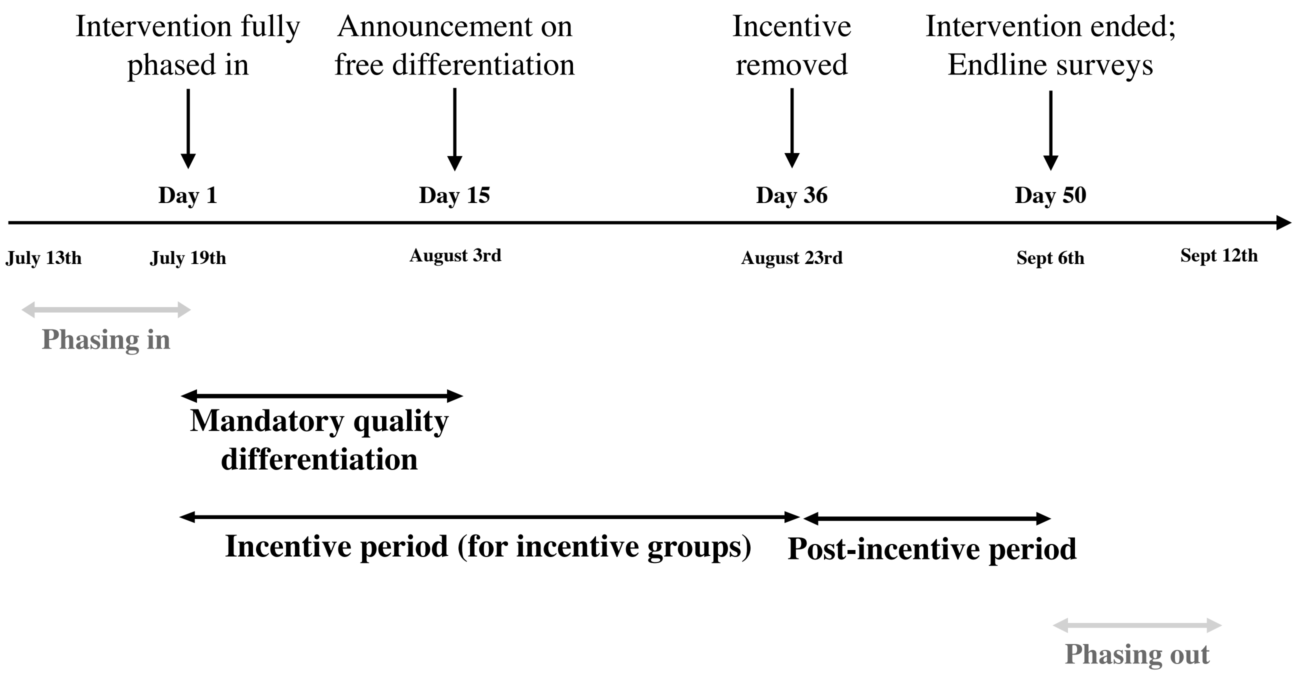
Panel B. The Sticker Group

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

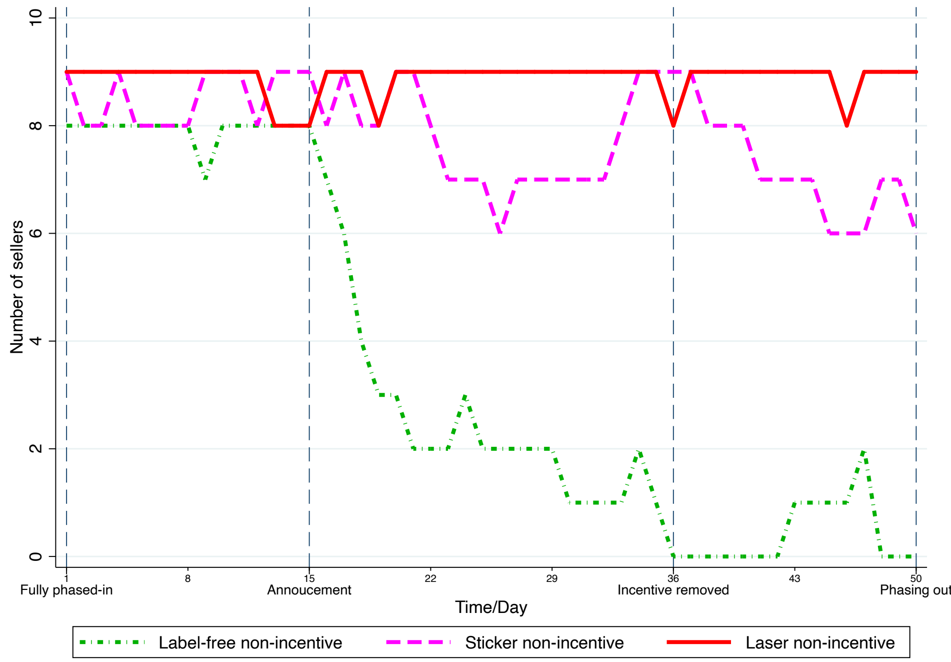


Note: This figure gives an overview of the timeline of the study.

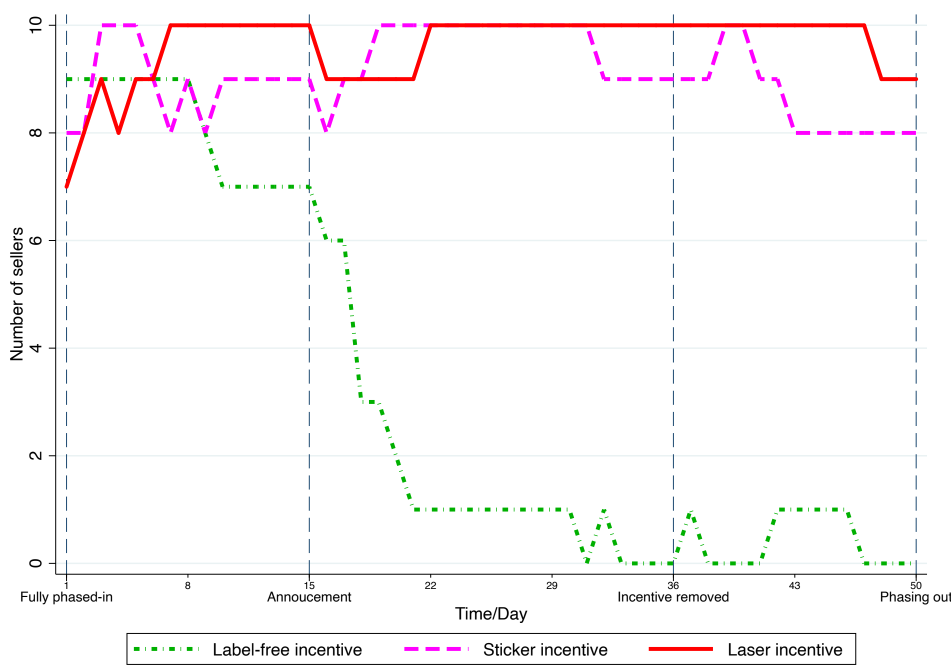
1. The intervention was rolled out from July 13 to 19, 2014.
2. All sellers were asked to experiment with quality differentiation for the first 2 weeks, from July 19 to August 3. To participate in the experiment, sellers signed an agreement form at the beginning of the period that they would experiment 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

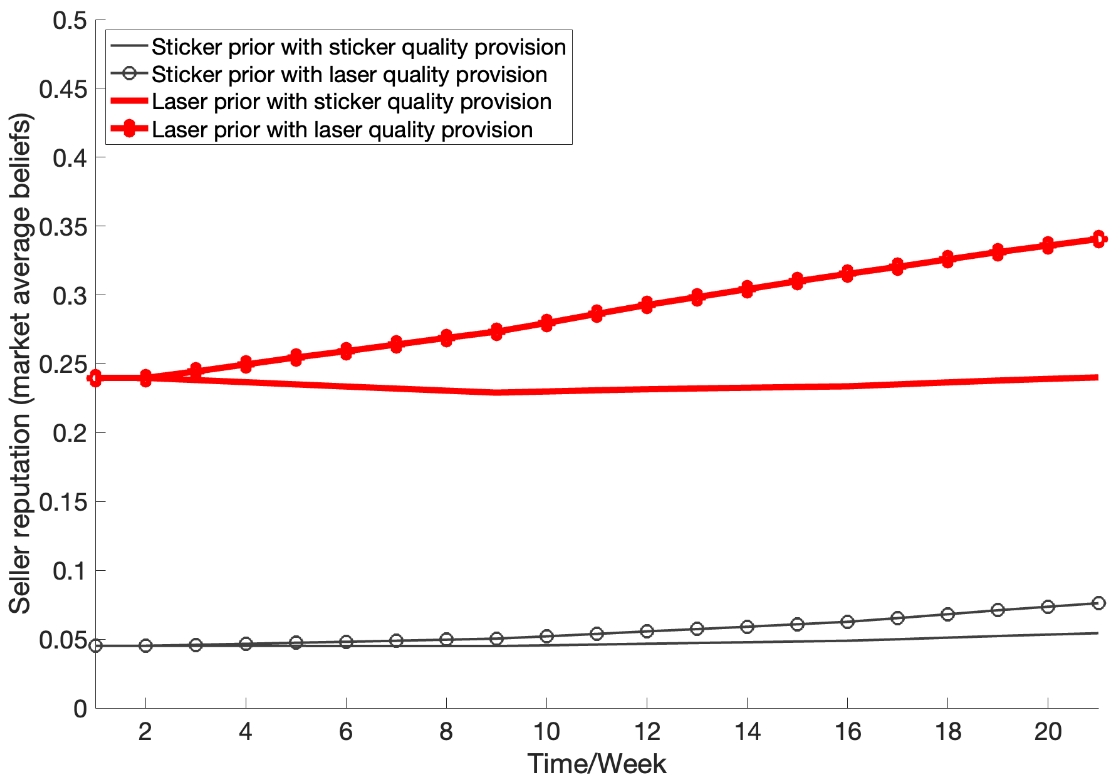


Panel B. Incentive Groups



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

Note: This table provides the summary statistics for sample characteristics of communities, sellers and households measured in the baseline surveys. In total, 60 sellers in 60 communities (markets) and 675 households were recruited for this study. Variation in the number of observations 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		✓		✓
Time fixed effects	✓	✓	✓	✓
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	0.02
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 effects	✓	✓	✓	✓
Time fixed effects	✓	✓	✓	✓
Normal pile mean	9.787	9.366	0.102	-0.285
Std. dev.	(0.99)	(0.923)	(0.774)	(0.965)
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**	0.045***	49.852*	49.454*	0.001	-0.001	-40.374	-55.550**	9.478	-6.096
	(0.199)	(0.196)	(0.015)	(0.015)	(28.758)	(28.506)	(0.010)	(0.009)	(24.860)	(23.831)	(39.378)	(41.676)
Laser	0.297*	0.396**	0.069***	0.065***	62.041***	70.450***	-0.006	-0.001	-12.445	-4.449	49.596	66.002**
	(0.154)	(0.156)	(0.020)	(0.019)	(22.073)	(23.296)	(0.010)	(0.010)	(26.705)	(18.699)	(36.728)	(31.906)
Observations	1452	1452	1456	1456	1462	1462	1456	1456	1462	1462	1462	1462
Baseline controls		✓		✓		✓		✓		✓		✓
Time fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Small sample robustness												
Permutation test (p-value):												
sticker	0.881	0.859	0.096	0.053	0.080	0.130	0.906	0.945	0.150	0.039	0.809	0.882
laser	0.132	0.068	0.000	0.007	0.0290	0.027	0.514	0.891	0.689	0.862	0.210	0.112
Clustered bootstrap (p-value):												
sticker	0.876	0.860	0.016	0.012	0.080	0.120	0.901	0.952	0.113	0.035	0.804	0.894
laser	0.061	0.026	0.001	0.003	0.006	0.010	0.528	0.895	0.659	0.835	0.188	0.078
Label-free Mean	4.284		0.055		56.313		0.011		180.475		236.788	
Std. dev.	(0.687)		(0.091)		(136.508)		(0.084)		(124.07)		(156.597)	

Note: This table examines sales profits, price and quantity for sellers in the non-incentive groups. Each observation is at the seller-day level. Sticker and laser are group dummies, and the omitted group is the label-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		✓		✓

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
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	✓		✓	
Time FE	✓	✓	✓	✓

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
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	✓	✓	✓	✓
Week Fixed Effects	✓	✓	✓	✓

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	0.105	0.045	-
	(0.007)	(0.003)	-
λ_l^0	0.268	0.240	-
	(0.135)	(0.003)	-
r	0.531	0.501	-
	(0.022)	(0.023)	-
α	0.174	0.135	0.192
	(0.088)	(0.001)	(0.001)
η	-2.667	-2.893	-2.697
	(0.695)	(0.004)	(0.008)
η_l	0.953	1.228	1.018
	(1.031)	(0.004)	(0.006)
τ_s	-1.907	-2.073	-1.961
	(0.037)	(0.006)	(0.003)
τ_l	-0.389	-0.186	-0.396
	(0.298)	(0.006)	(0.002)
Market FE	✓	✓	✓
Time FE	✓	✓	✓
Log-likelihood	-3051.470	-1122.995	-1199.637

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 5 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.

Appendices

A Tables and Figures

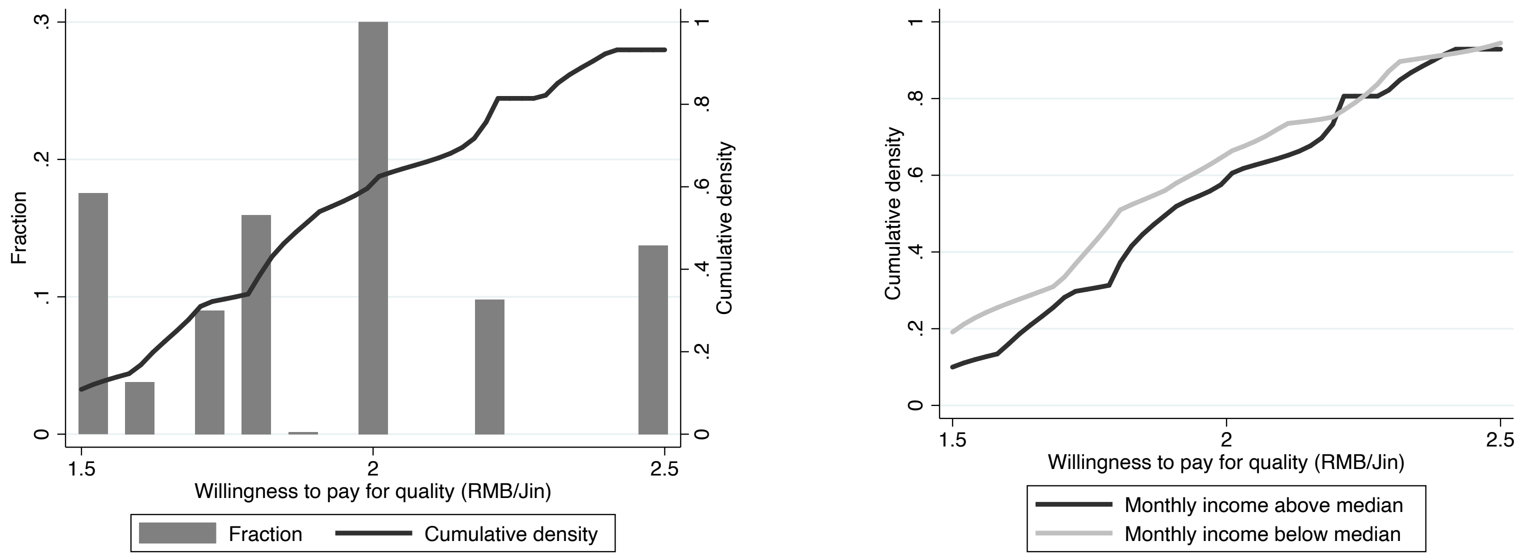
Figure A.1: A Local Market

Note: This figure shows a picture of a typical local market in Shijiazhuang, China.

Figure A.2: A Sweetness Meter

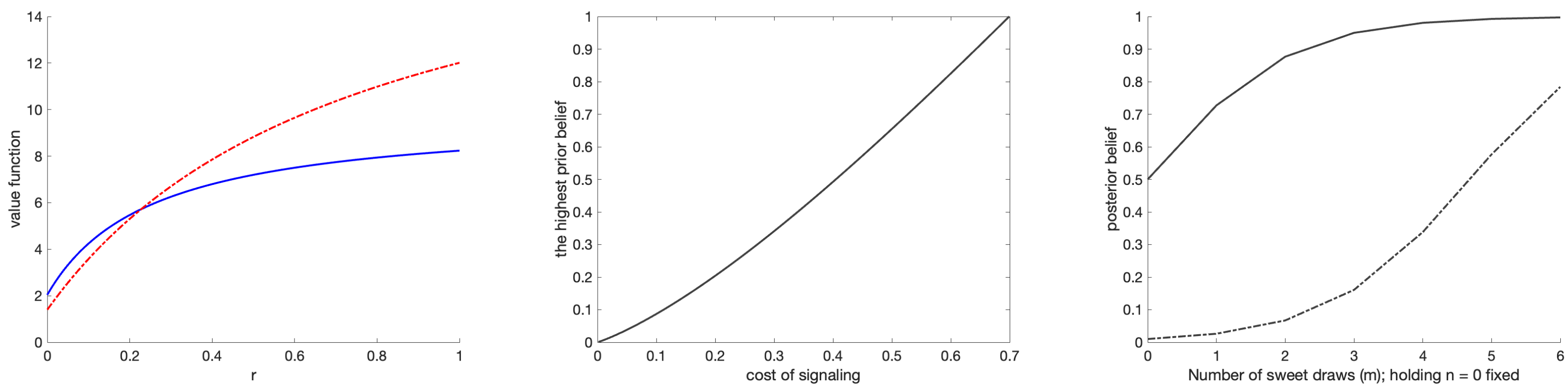
Note: This figure shows a picture of a sweetness meter, which reports the Brix degree.

Figure A.3: Consumers' Self-reported Willingness to Pay for Quality



Note: This figure shows the heterogeneity of households' self-reported willingness to pay for quality elicited in the baseline survey. Households were asked to consider a hypothetical situation where they see two piles of watermelons sold in the local market, one pile of ordinary quality at 1.5 RMB/jin and the other pile of premium quality but at a higher price. Surveyors announced the price for the premium pile from high to low and recorded the highest number that led to the choice of the premium pile. The sequence of prices (in RMB/jin) was announced in the following order: 2.5, 2.2, 2, 1.9, 1.8, 1.7, 1.6, 1.5. The left figure plots the empirical distributions of willingness to pay for quality for all households in the baseline survey; the right figure groups household by monthly income above and below the median.

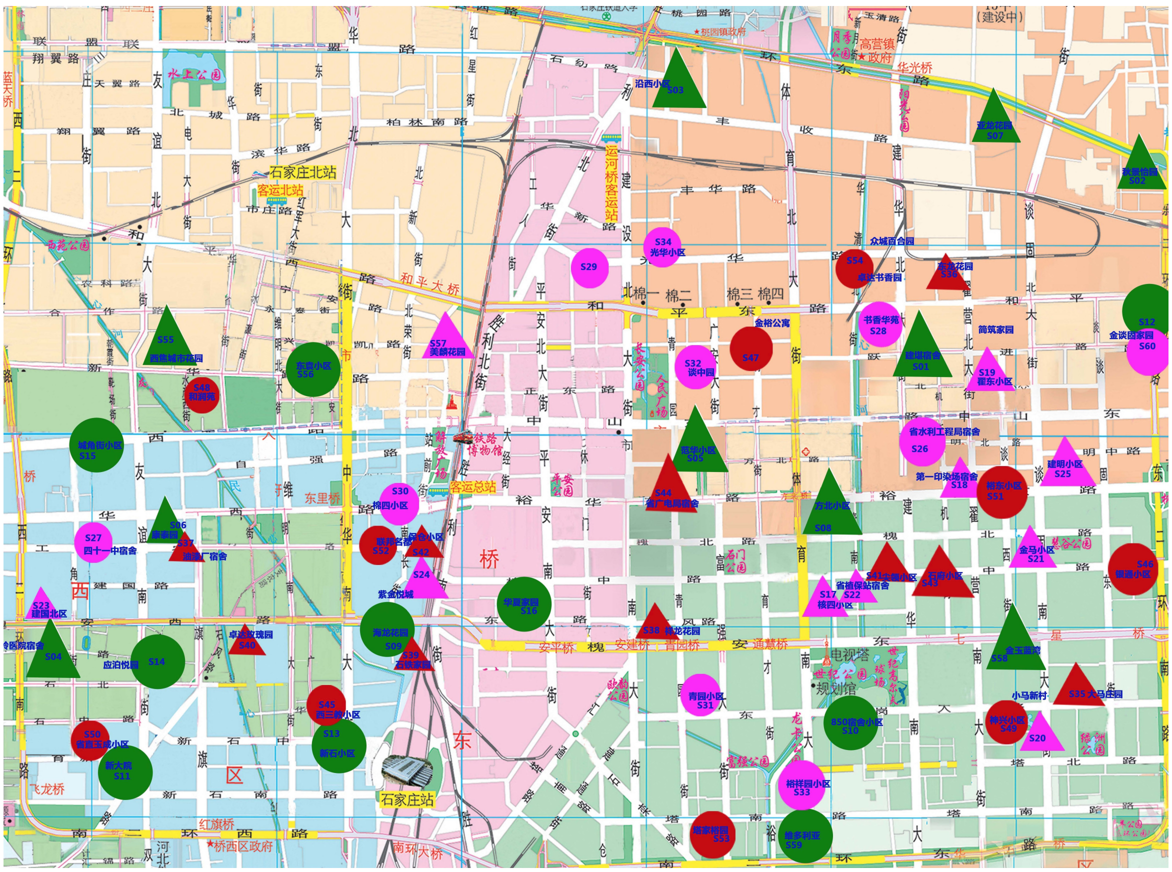
Figure A.4: Theoretical Model Simulation: Equilibrium Analysis



(a) (b) (c)

Note: Panel (a) simulates the full separating equilibrium in Section 3.3 under the following set of parameter values: $c = 0.1, \underline{\gamma} = 0.3, \overline{\gamma} = 0.8, \lambda_0 = 0.5, \delta = 0.95$ and different values of r. The solid line plots V_G^+ against different values of r, and the dotted line plots W_G^+. Panel (b) plots the highest prior belief (the largest λ_0) that can be supported in a partial pooling equilibrium (as discussed in Section 3.4) against the cost of signal. Panel (c) plots $\tilde{\lambda}$ against the number of sweet draws m, holding the number of non-sweet draws n to be 0. The solid line assumes $\lambda_0 = 0.5$ and the dotted line assume $\lambda_0 = 0.01$.

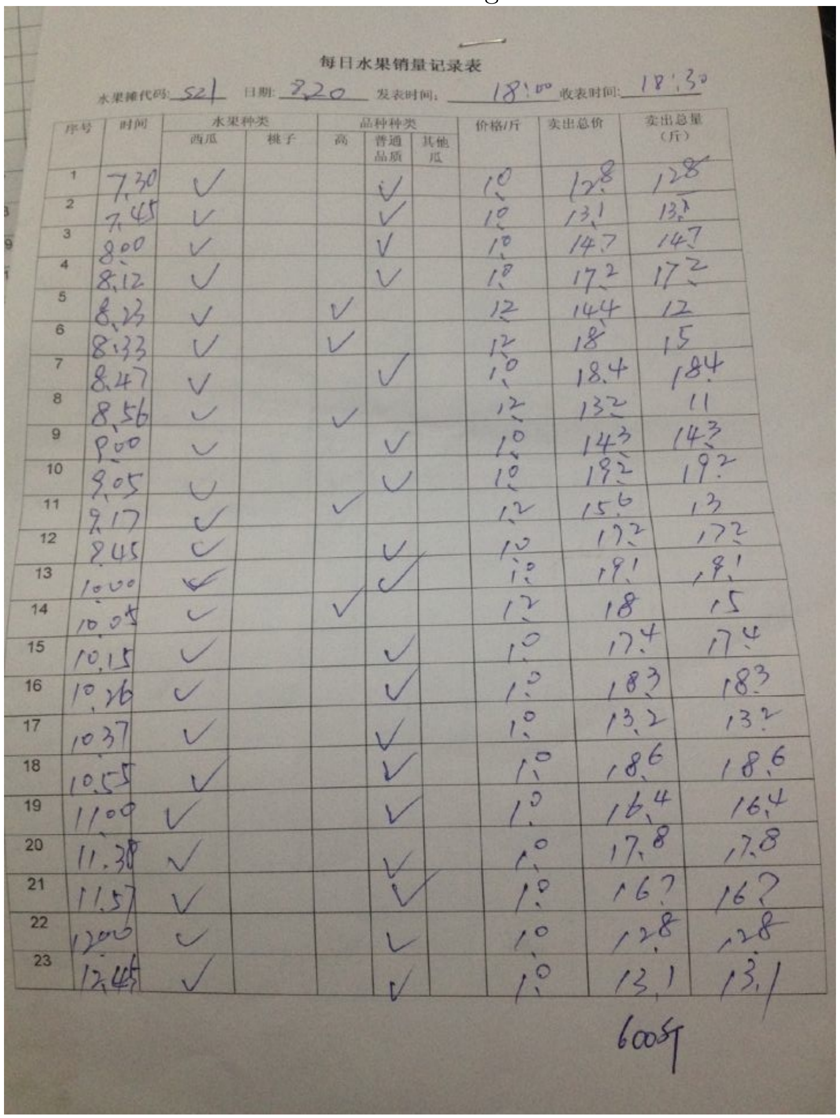
Figure A.5: Map of the Randomization



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

Figure A.6: Examples of Seller and Household Recording Sheets

A. Seller Recording Sheet



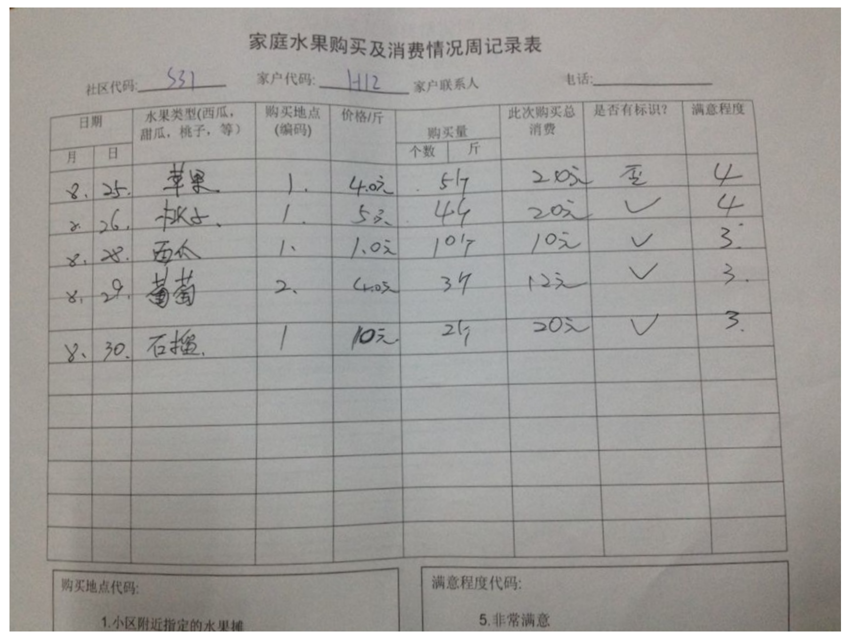
每日水果销量记录表

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

序号	时间	水果种类 西瓜	水果种类 桃子	品种种类 高	品种种类 普通品质	品种种类 其他瓜	价格/斤	卖出总价	卖出总量（斤）
1	7.30	✓			✓		1.0	12.8	12.8
2	7.45	✓			✓		1.0	13.1	13.1
3	8.00	✓			✓		1.0	14.7	14.7
4	8.12	✓			✓		1.0	17.2	17.2
5	8.23	✓		✓			1.2	14.4	12
6	8.33	✓		✓			1.2	18	15
7	8.47	✓			✓		1.0	18.4	18.4
8	8.56	✓		✓			1.2	13.2	11
9	9.00	✓			✓		1.0	14.3	14.3
10	9.05	✓			✓		1.0	19.2	19.2
11	9.17	✓		✓			1.2	15.6	13
12	9.45	✓			✓		1.0	17.2	17.2
13	10.00	✓			✓		1.0	18.1	18.1
14	10.05	✓		✓			1.2	18	15
15	10.15	✓			✓		1.0	17.4	17.4
16	10.26	✓			✓		1.0	18.3	18.3
17	10.37	✓			✓		1.0	13.2	13.2
18	10.55	✓			✓		1.0	18.6	18.6
19	11.00	✓			✓		1.0	16.4	16.4
20	11.38	✓			✓		1.0	17.8	17.8
21	11.57	✓			✓		1.0	16.7	16.7
22	12.00	✓			✓		1.0	12.8	12.8
23	12.45	✓			✓		1.0	13.1	13.1

600斤

B. Household Recording Sheet



家庭水果购买及消费情况周记录表

社区代码: S31 家户代码: H12 家户联系人 电话:

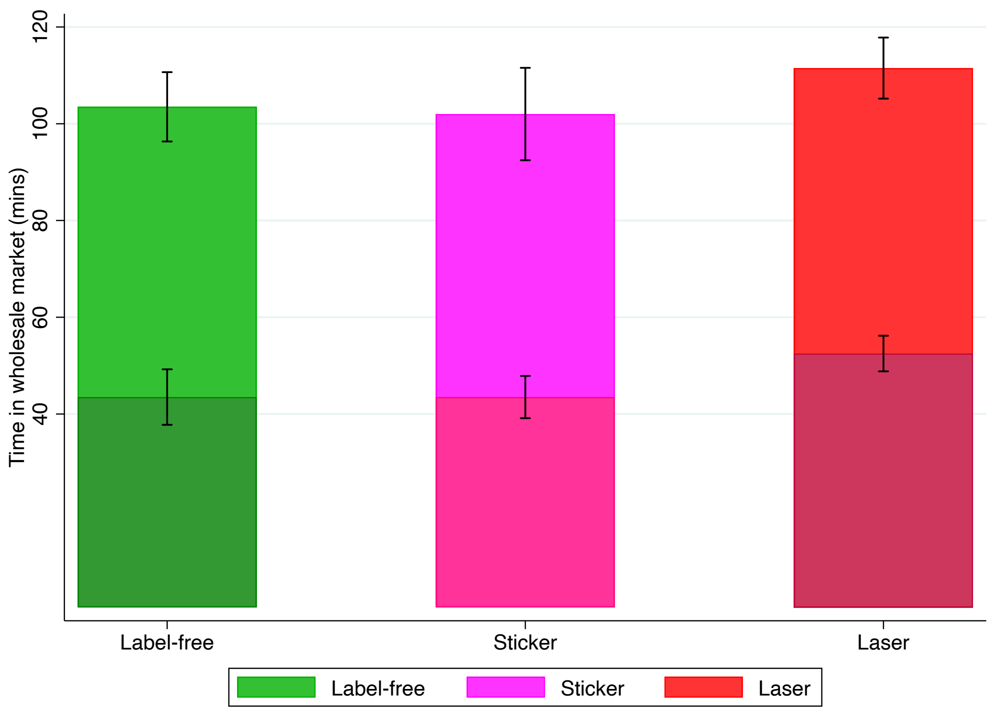
日期 月	日期 日	水果类型(西瓜，甜瓜，桃子，等)	购买地点(编码)	价格/斤	购买量 个数	购买量 斤	此次购买总消费	是否有标识?	满意程度
8.	25.	苹果	1.	4.0元		5斤	20元	否	4
8.	26.	桃子	1.	5元		4斤	20元	✓	4
8.	28.	西瓜	1.	1.0元		10斤	10元	✓	3.
8.	29.	葡萄	2.	4.0元		3斤	12元	✓	3.
8.	30.	石榴	1	10元		2斤	20元	✓	3.

购买地点代码:
1.小区附近指定的水果摊

满意程度代码:
5.非常满意

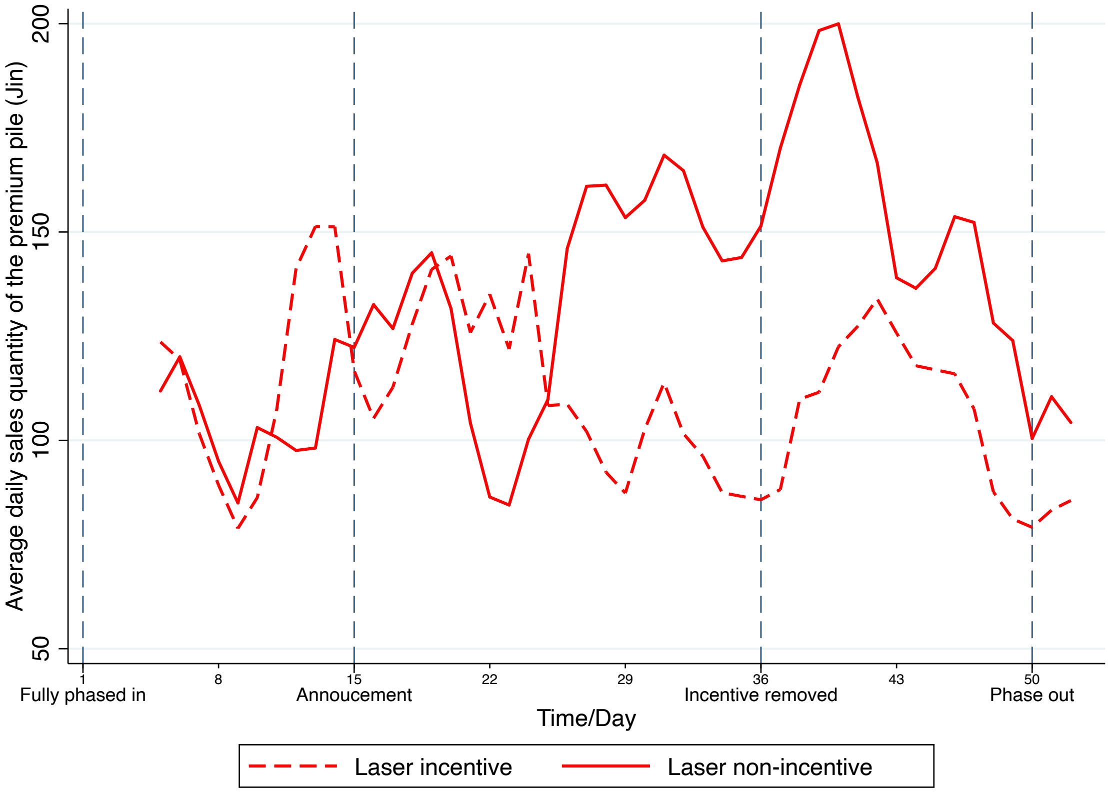
Note: This figure shows an example of a seller recording sheet (daily) and a household recording sheet (weekly).

Figure A.7: Sourcing Efforts in Wholesale Markets



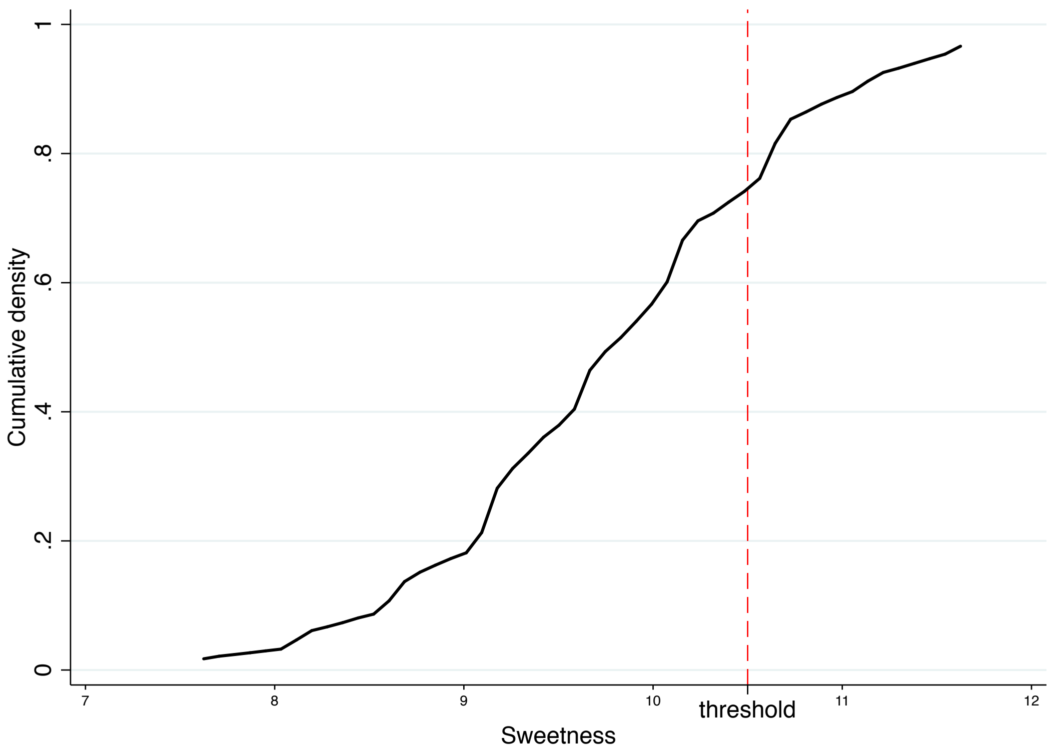
Note: This figure compares sellers' self-reported behavior in the wholesale market. It plots the mean and standard deviation for total time spent in the wholesale market (lighter color) and time spent on sourcing watermelons (darker color).

Figure A.8: Sales Dynamics: Laser Incentive vs. Laser Non-Incentive



Note: This figure plots the average daily sales of the premium watermelons for the laser groups. The solid line plots the daily average for the laser incentive group, and the dashed line plots that for the laser non-incentive group.

Figure A.9: Distribution of Sweetness for the Undifferentiated Pile



Note: This figure plots the empirical cumulative distribution of the measured sweetness for the label-free group after the sellers reverted to non-differentiation. The threshold, 10.5, marks the criterion for receiving the incentive, corresponding to 73%.

Table A.1: Balance Check for Baseline Community and Market Characteristics

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

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

Table A.2: Balance Check for Baseline Seller Characteristics

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

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

Table A.3: Balance Check for Baseline Household Characteristics

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

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

Table A.4: Effects on Quality Provision Using Household Satisfaction Ratings

	Ordered probit: Satisfaction rating from 1 to 5		Probit: Dummy for the highest rating of 5	
	All	Non-incentive Only	All	Non-incentive Only
	(1)	(2)	(3)	(4)
Laser	0.553***	0.496**	0.469**	0.482*
	(0.176)	(0.241)	(0.206)	(0.292)
Observations	258	127	258	127
Time Fixed Effects	✓	✓	✓	✓

Note: This table examines quality provision by labeling treatment groups using the data collected from households. Quality is measured in terms of the satisfaction rating for watermelons purchased from the premium pile. Each observation is a household-week purchase. Columns 1 and 2 estimate an ordered probit model using the original self-reported satisfaction rating ranging from 1 to 5. Columns 3 and 4 estimate a probit model for a dummy variable for the highest satisfaction rating of 5. All regressions include week fixed effects. Standard errors are clustered at the household level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table A.5: Pricing Behavior of Competitors

Dep. var.: Daily market average price charged by competitors (in RMB/jin)

	All	Non-incentive Only
	(1)	(2)
Sticker	-0.014	-0.007
	(0.009)	(0.009)
Laser	0.006	0.015
	(0.014)	(0.011)
Observations	1462	2900
Omitted group mean	0.964	0.963
Std. dev.	(0.072)	(0.063)
Day fixed effects	✓	✓

Note: This table examines the pricing behavior of other sellers operating in the markets who are not included in the study sample across treatment groups. The dependent variable is the daily market average price charged by other sellers (measured in RMB/jin). The omitted group is the label-free group. Column 1 include all sellers, and Column 2 considers only those in the non-incentive group. Standard errors are in parentheses and clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table A.6: Effects of Labeling and Incentive Treatments on Quality Provision

Dep. var.: Quality of premium pile (measured in sweetness)				
	All weeks		First 5 weeks	
	(1)	(2)	(3)	(4)
Laser non-incentive	0.712***	0.638***	0.747***	0.630**
	(0.217)	(0.228)	(0.252)	(0.262)
Laser incentive	1.205***	1.154***	1.242***	1.155***
	(0.162)	(0.170)	(0.193)	(0.197)
Sticker incentive	0.919***	0.882***	1.031***	0.959***
	(0.131)	(0.133)	(0.171)	(0.163)
Observations	468	468	391	391
Baseline controls		✓		✓
Time fixed effects	✓	✓	✓	✓
Sticker non-incentive mean	9.738			
Std. dev.	(1.104)			
Clustered bootstrap (p-value):				
Laser non-incentive	0.002	0.015	0.004	0.036
Laser incentive	0.000	0.000	0.000	0.000
Sticker incentive	0.000	0.000	0.000	0.000

Note: This table examines quality provision by treatment group. Quality is measured in sweetness. Each observation is at the seller-check level. The sample includes all sticker and laser markets. Columns 1 and 2 include all weeks, and Columns 3 and 4 restrict the sample to the first 5 weeks prior to the removal of the incentive. 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. Standard errors are clustered at the seller level. Clustered bootstrap is used to perform nonparametric bootstrap estimation of the regression coefficients (1,000 replications). ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table A.7: Heterogeneity in Sellers' Sorting Ability

Dep. var.: ability (dummy)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
gender (male=1)	0.333**							0.245
	(0.147)							(0.145)
age		0.006						0.003
		(0.008)						(0.009)
years of experience			0.034***					0.025*
			(0.012)					(0.013)
years of schooling				-0.013				-0.004
				(0.027)				(0.027)
log housing price					-0.331			-0.297
					(0.246)			(0.224)
log number of housing units							-0.114	-0.132
							(0.103)	(0.095)
Observations	30	30	30	30	30	30	30	30

Note: This table examines the heterogeneity in sorting ability among sellers at baseline using the data from the sorting test. Ability is a dummy variable that equals to 1 if and only if the seller did not make any *clear mistake* in the sorting test. A 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. Among 30 sellers participated in the sorting test, 7 made any clear mistake.

Table A.8: Household Purchasing Dynamics: Laser Non-incentive vs. Incentive Markets

	Laser Non-incentive Markets		Laser Incentive Markets	
	(1)	(2)	(3)	(4)
Lagged avg. satisfaction rating	0.251***		0.315**	
	(0.085)		(0.116)	
Lagged % of very good experiences		0.330***		0.399**
		(0.096)		(0.148)
Observations	103	105	88	88
Household baseline controls	✓	✓	✓	✓
Week fixed effects	✓	✓	✓	✓

Note: This table examines the purchasing dynamics between laser-incentive and laser non-incentive markets. Each observation is at the household-week level. The dependent variable is whether the household has purchased any watermelons from the premium pile in a given week. The key regressors are two measures of lagged experiences: (1) the average lagged satisfaction rating (ranging from 1 to 5) of all premium watermelons purchased prior to the period; (2) the percentage of past consumption experiences that attained the highest satisfaction rating of 5. All regressions include week fixed effects and controls for the following set of household baseline characteristics: household size, percentage elderly residents, monthly income, average number of watermelons consumed per week reported in the baseline survey, and the baseline self-reported WTP 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 A.9: Spillover Effect to Peach Sales

Dep var: daily peach sales (in RMB)

	(1)	(2)
Label-free incentive	-14.029	-23.361
	(29.712)	(28.593)
Sticker non-incentive	30.190	41.283
	(37.136)	(31.877)
Sticker incentive	-17.491	-23.835
	(29.814)	(28.966)
Laser non-incentive	-13.605	-30.616
	(28.715)	(29.367)
Laser incentive	42.153	29.056
	(36.721)	(34.705)
Constant	113.354***	9.557
	(26.722)	(61.867)
Observations	1312	1312
Time fixed effects	✓	✓
Seller (Market) baseline controls		✓

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

Table A.10: Time Dynamics for Sales of the Premium Pile

	Laser		Sticker	
	(1)	(2)	(3)	(4)
Day	-0.576		-0.508	
	(0.385)		(0.803)	
DayXIncentive	1.598***		-0.309	
	(0.494)		(0.903)	
Week		-3.405		-3.589
		(2.635)		(5.670)
WeekXIncentive		11.367***		-1.512
		(3.432)		(6.377)
Control Mean	117.509***	111.706***	129.573***	129.656***
	(7.297)	(7.384)	(13.783)	(13.681)
Observations	971	971	976	976
Seller fixed effects	✓	✓	✓	✓

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

Table A.11: Household Purchasing Probabilities Conditional on Experiences

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

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

Table A.12: Simulated Maximum Likelihood Estimation: Robustness Checks

Parameters	$\overline{\gamma} = 0.625$ (1)	$\overline{\gamma} = 0.75$ (2)	$\overline{\gamma} = 0.5$ (3)	No Market FEs (4)
λ_s^0	0.045	0.000	0.079	0.086
	(0.003)	(0.008)	(0.001)	(0.231)
λ_l^0	0.240	0.218	0.274	0.247
	(0.003)	(0.003)	(0.001)	(0.680)
r	0.501	0.500	0.501	0.548
	(0.023)	(0.023)	(0.024)	(0.045)
α	0.135	0.136	0.135	0.192
	(0.001)	(0.001)	(0.000)	(0.028)
η	-2.893	-2.878	-2.894	-2.724
	(0.004)	(0.011)	(0.002)	(0.292)
η_l	1.228	1.205	1.251	0.969
	(0.004)	(0.005)	(0.002)	(0.504)
τ_s	-2.073	-2.073	-2.073	-1.961
	(0.006)	(0.022)	(0.002)	(0.003)
τ_l	-0.186	-0.186	-0.186	-0.396
	(0.006)	(0.005)	(0.002)	(0.231)
Market FE	✓	✓	✓	
Time FE	✓	✓	✓	✓

Note: This table examines robustness of the simulated maximum likelihood estimation results using the sub-sample of households with more than 5 purchases during the season. Column 1 reproduces the baseline estimates in Column 2 of Table 7. Columns 2 and 3 consider alternative values of $\overline{\gamma}$. Column 4 excludes market fixed effects. Standard errors shown in parentheses are calculated as the square root of the inverse of the Hessian matrix.

B Additional Tests, Sample Recruitment and Data

B.1 Sorting Ability Test

To establish information asymmetry in the market, a sorting ability test was conducted with 30 fruit sellers from 30 different local markets in the city, representing half of the experimental sample. Additionally, the same test was repeated with 5 randomly chosen local consumers in each market. During the test, each participant was asked to sort 10 watermelons into two piles: one for high quality and one for low quality. To avoid influencing the results, enumerators did not specify a fixed number of watermelons for each pile but suggested a range, with the maximum and minimum for each pile set at 7 and 3, respectively. On average, sellers sorted 4.4 watermelons into the premium pile, while consumers sorted 3.5. The watermelons used in the test were randomly selected by surveyors from the sellers' stores, with no obvious distinguishable differences in appearance. Finally, to measure quality, a sweetness meter was utilized.

B.2 Pre-Intervention Survey: Willingness to Pay under Different Labels

A pre-intervention survey with 300 consumers in the city of Shijiazhuang was conducted, during which consumers were asked about their willingness to pay (WTP) for watermelons sold with different labels (without further information on quality). Specifically, the following questions were asked:

1. Suppose that it is a summer day. You go to the local fruit seller near your residential estate to buy watermelons. You see two piles of watermelons. [Instruction to surveyor: Show the respondent two photos of watermelons, one unlabeled and one sticker-labeled with the words "Premium Watermelons" ("Jing Pin Xi Gua").] The first pile (unlabeled) is sold at 1.5RMB/jin in the market. The sticker-labeled pile is sold at XX RMB/jin. Without any further information about quality, which one would you buy? [Instruction to surveyor: Read out the hypothetical prices sequentially from high to low (2RMB, 1.9 RMB, 1.8 RMB, 1.7 RMB, 1.6 RMB, 1.5 RMB), and record the highest price that the respondent is willing to pay for the high quality type.]

2. Now suppose that you see another two piles of watermelons. [Instruction to surveyor: Show the respondent two photos of watermelons, one unlabeled and one laser-labeled with the words "Premium Watermelons" ("Jing Pin Xi Gua").] Again, the first pile (unlabeled) is sold at 1.5RMB/jin in the market. The laser-labeled pile is sold at XX RMB/jin. Without any further information about quality, which one would you buy? [Instruction to surveyor: Read out the hypothetical prices sequentially from high to low (2RMB, 1.9 RMB, 1.8 RMB, 1.7 RMB, 1.6 RMB, 1.5 RMB), and record the highest price that the respondent is willing to pay for the high-quality type.]

B.3 Experimental Sample Recruitment

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

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

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

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

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

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

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

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

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

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

B.4 Seller Data Cleaning

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

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

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

B.5 Household Data Cleaning

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

Broadly speaking, there are two issues with the household data. First, for some transaction records, one or more of the following information may be missing: the date of purchase, fruit type, purchase place, purchase quantity, purchase value, whether the fruit purchased has any labels, and the self-reported satisfaction rating. Second, the records are missing for some households in some weeks, due to either family travels or other obstacles that led to failure of collection. The latter is less of an issue than the former since weekly fruit coupons were distributed upon collecting the recording sheets, which gave households an incentive to turn in their forms on time.

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

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

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

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

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

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

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

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

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

The final analysis sample contains no missing place information and no missing pile information for purchases made at the sample seller. The satisfaction rating is missing for 11% of the purchases. Overall, 30.7% of the recorded purchases are made from the sample sellers' stores, and 57.2% are made from other sellers located in the same local market and nearby supermarkets; the rest are from other places. On average, each household buys 1.1 watermelons per week, and the median is 1. These descriptive patterns all look similar to those for the full sample.

B.6 Household Endline Survey: Willingness to Pay under Different Labels

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