Melons as Lemons:

Asymmetric Information, Consumer Learning and Seller Reputation

Jie Bai *

August 2016

JOB MARKET PAPER

Abstract

There is often a lack of reliable high quality provision in many markets in developing countries. I designed an experiment to understand this phenomenon in a setting that features typical market conditions in a developing country: the retail watermelon market in a major Chinese city. I begin by demonstrating empirically that there is substantial asymmetric information between sellers and buyers on sweetness, the key indicator of quality for watermelons, yet sellers do not sort and price watermelons by quality. I then randomly introduce one of two branding technologies into 40 out of 60 markets—one sticker label that is widely used and often counterfeited and one novel laser-cut label. I track sellers' quality, pricing and sales over an entire season and collect household panel purchasing data to examine the demand side's response. I find that laser branding induced sellers to provide higher quality and led to higher sales profits, establishing that reputational incentives are present and can be made to pay. However, after the intervention was withdrawn, all markets reverted back to baseline. To rationalize the experimental findings, I build an empirical model of consumer learning and seller reputation. The structural estimates suggest that consumers are hesitant to upgrade their perception about quality under the existing branding technology, which makes reputation building a low return investment. While the new technology enhances consumer learning, the resulting increase in profits is not sufficient to cover the fixed cost of the technology for small individual sellers. Counterfactual analysis shows that information frictions and fragmented markets lead to significant under-provision of quality. Third-party interventions that subsidize initial reputation building for sellers could improve welfare.

*Department of Economics, Massachusetts Institute of Technology. E-mail: jieb@mit.edu. I am extremely grateful to my advisors Benjamin Olken, Abhijit Banerjee, Robert Townsend and Nikhil Agarwal for their continuous encouragement and guidance over the course of this project. I also thank David Atkin, Esther Duflo, Glenn Ellison, Robert Gibbons, German Sergio Gieczewski, Seema Jayachandran, Yan Ji, Siyuan Liu, Rocco Macchiavello, Yuhei Miyauchi, Xiaosheng Mu, Ariel Pakes, Parag Pathak, Frank Schilbach, Peng Shi, Maheshwor Shrestha, Paulo Somaini, Tavneet Suri, Jean Tirole, Juuso Toikka, Richard Zeckhauser, Juanjuan Zhang, Hongkai Zhang and participants of the development seminar, development lunch, and IO lunch at MIT for helpful discussions and feedback. Funding for this project was generously provided by the Weiss Family Fund, the Abdul Latif Jameel Poverty Action Lab, and the George and Obie Shultz Fund. All errors are my own.

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 for experience goods, such as food products and drugs. For example, in China, there are rising public concerns over food quality and safety in recent years. This leads policy makers and academics to question whether the problem is due to low demand for quality–as a result of lower income and higher cost of production compared to the developed world–or due to other reasons. It is well recognized in economics that when contracting on quality is difficult information frictions can lead to quality deterioration and firms typically need a good reputation to succeed. However, a reputation for quality is precisely what many firms in developing countries are lacking. In fact, a major challenge faced by these firms is the difficulty to move up the value chain into producing higher quality and higher value-added products. The question is, then, what are the main barriers that prevent them from doing so? Why is there a lack of premium quality brands in many markets? Answering this question can have important policy implications: first, on a broad level, it helps governments to structure policies to facilitate firms' quality upgrading; second, in the context of food products and drugs, it may offer new solutions to address the information problem, alternative to direct government regulations and quality controls, which could be very costly to enforce especially in countries with weak legal systems.

In this paper, I designed an experiment to understand what hinders quality provision in a developing country setting: the retail watermelon markets in a major Chinese city. The environment has many features in common with markets for daily goods in developing countries. Several features of this market make it particularly suited for studying this topic. First, there are a large number of small independent local markets, which allows randomization at the market level. These markets are also highly unregulated at baseline, thus one could imagine that the one mechanism that could potentially work is the reputation mechanism. Therefore, the setting provides a clean environment to study seller reputation. Second, the quality of a watermelon can be captured very well by its sweetness, which can be measured (ex post) using a sweet meter. This maps nicely into a one-dimension vertical differentiation model. Third, I use survey data to document that there is substantial asymmetric information about quality between sellers and buyers. While consumers find it difficult to detect quality at the point of sale, sellers have the ability to assess quality based on less obvious observables. However, despite this and despite considerable variation in quality within batches of watermelons, there is a stark absence of quality differentiation at baseline. Sellers sell an undifferentiated pile of watermelons and charge a uniform price, and there is also little price variation across sellers in one market. The goal of this research is to first examine if the outcome is efficient–that is, there is no market failure given consumers' valuation for quality, the underlying cost of differentiation and the current market structure; if not, why that is the case and what it takes to motivate quality provision. The lack of quality differentiation is not unique to the watermelon market. Exploiting features of the watermelon markets, this study seeks to shed light on sellers' and buyers' behavior under asymmetric information and provide a framework for thinking about quality provision in other similar settings.

I first propose a model for thinking about quality provision in this setting. The model features a long-run seller who chooses quality to maximize the expected discounted sum of profits, subject to a dynamic demand system that is rooted in consumer learning. It highlights two broad explanations for the lack of quality differentiation at baseline. First, it could be that the cost of reliably providing high quality is too high relative to consumers' valuation for quality. Consequently, higher quality is neither demanded nor supplied. Second, due to the information problem, a seller's claim of offering high quality cannot be immediately verified. Since reputation building takes time, consumers' initial perception and speed of learning matter for sellers' reputational incentive. In particular, pessimistic beliefs could make reputation building a low return investment. Sellers who rationally discount future profits may lack the incentive to build a reputation for quality. In this situation, markets are stuck in an equilibrium with no quality differentiation. The welfare and policy implications of these two explanations differ: in the first case, the distortion caused by the information problem is small, while in the second, it could be large. The model predicts that enhancing consumers' prior beliefs could potentially strengthen sellers' reputational incentive and motivate higher quality provision.

Therefore, to understand what hinders quality differentiation at baseline, I conducted a field experiment with 60 sellers located in 60 different markets in Shijiazhuang, China. I randomly introduced one of two branding technologies into 40 out of the 60 markets—one sticker label that is widely used and often counterfeited, and one novel laser-cut label. Consumer pilot survey shows that consumers believe that the laser branding is more effective at deterring counterfeiting activities because laser machines are very expensive. The new branding technology could potentially dispel negative historical stereotypes, thereby allowing sellers to establish trust faster. The model suggests that sellers in the laser group may have a stronger incentive to provide quality. Next, for a cross-randomized subset of sellers, I further provided them with a temporary monetary incentive to invest in their new high-quality brand. The incentive treatment provides a direct test of the model's predictions: if the incentive facilitates sellers' initial reputation building, then upon its removal sellers who have had the incentive would be endowed with a higher reputation than those who have not. The model suggests that higher quality may sustain even in the post-incentive period.

The intervention lasted over eight weeks, spanning the entire peak season for watermelons. To avoid spillovers across different branding treatments, only one seller was sampled in each market. Each of the 60 sellers was asked to sell two piles of watermelons at the retail site: a premium pile and a normal pile. Sellers were free to set the quality, price, and quantity for each pile. They were then randomized into three branding treatments for their watermelons in the premium pile: a laser engraving of the words "premium watermelon"; a sticker label with the same words; or no labeling. Quality differentiation was mandatory for the first two weeks but sellers were free to decide afterwards. This was designed in order to examine the differential incentives across the branding groups. The incentive treatment was enforced through biweekly quality checks, and was lifted at the end of the sixth week (unanticipated by sellers).

Data on pricing and sales were collected from sellers' daily sales records and surveyors' daily market visits. Quality was measured from the biweekly quality checks using sweet meters. To examine the demand side's response to quality differentiation, 675 households from 27 markets, evenly distributed across the treatment groups, were recruited to record the family's summer fruit purchasing and consumption. Endline and follow-up surveys were conducted to elicit changes in consumers' perceptions and sellers' longer-term behavior.

First, both laser and sticker branding induced sellers to differentiate quality beyond the mandatory period, whereas sellers in the label-less group sharply reverted back to no differentiation after the first two weeks. Conditional on differentiation, sellers in the laser group provided a genuine quality-price premium, establishing that reputational incentives are present and could discipline sellers' behavior. On the other hand, evidence for the sticker group is quite mixed: on average, quality of the premium pile was not significantly higher than the market average. Next, the incentive treatment successfully induced sellers in both sticker and laser groups to provide higher quality than their non-incentivized counterparts. However, higher quality was only sustained for the laser incentive group. Sellers in the sticker incentive group reverted to a lower quality level after the incentive was removed. This is consistent with the previous discussion that it may take a long time to establish trust under the existing "contaminated" branding technology.

Overall, the experimental evidence is consistent with the model's predictions and establishes that reputational forces are at work. To understand why sellers who were induced to differentiate quality under the experiment were not already doing so at baseline, I exploit the fact that most sellers in the label-less group reverted back to no differentiation after the first two weeks, and compare the sales outcomes across different branding groups. I find that quality differentiation under sticker did not outperform no differentiation. In contrast, sellers in the laser group earned 30-40% higher sales profits on average than those in the label-less group. The increase can be attributed to both higher prices of the premium product and higher sales as a result of attracting more high-end consumers over time. These results demonstrate that there is a demand for quality and that reputation can be made to pay. Having said that, one year after the intervention when the laser branding technology was no longer available for free, all sellers reverted back to baseline. This suggests that individual sellers may not have the incentive to invest in the new technology themselves.

The experimental findings provide a qualitative explanation for the lack of quality differentiation at baseline. Next, I build an empirical model of consumer learning and seller reputation to rationalize the observed behavior and explain the experimental results.

The structural estimation proceeds in two steps. In the first step, I estimate a dynamic discrete choice demand system that explicitly models consumers' learning process and prior beliefs, which are allowed to vary under different branding technologies. The demand model incorporates rich heterogeneity in household preferences, and controls for a full set of market and time fixed effects to correct for potential price endogeneity. The model is estimated using simulated maximum likelihood and I exploit purchasing patterns and experience realizations observed in the household panel data for identification. In the second step, I solve numerically for the seller's optimal pricing and quality, taking the demand model as estimated, and apply a minimum distance estimator to recover the seller's discount factor and unobserved effort costs from observed empirical policies. To deal with the computational challenge that the state space in the seller's optimization problem is of infinite dimension (i.e. the joint distribution of household characteristics and beliefs), I restrict attention to the class of once-for-all quality and markup policies. The empirical patterns in the data provide some qualitative justification for this assumption. That being said, the data, which lasts only for eight weeks, is limited in investigating longer-term reputation dynamics. Therefore, one needs to be careful in interpreting the counterfactual results. Finally, in order to match the experimental setting, I work with a model of monopolistic sellers, but also perform several counterfactual exercises to investigate the effects of market competition.

The structural estimates describe the data well. Purchasing patterns generated by the Bayesian learning process fit actual purchasing behavior. The profit-maximizing markup for selling an undifferentiated pile of watermelons, as implied by the estimated demand elasticity, closely matches the empirical markup. Finally, the simulated sales outcomes under the dynamic demand system and the empirical policies mimic the actual sales in the seller data.

The structural estimates indicate that consumers' prior beliefs are more "stubborn" under sticker than under laser. As a result, trust can take a long time to establish, which explains why sellers do not have the incentive to build reputation under the existing signaling technology. While the new technology enhances consumer learning and thereby strengthens sellers' reputational incentive, the increase in the discounted return from quality differentiation, taking into account effort costs, is not large enough to justify the fixed cost of the technology for individual sellers. This explains why all sellers reverted back to baseline after the intervention was withdrawn. There could be two reasons: (a) the sellers' market size is small; and (b) it may be difficult for sellers to extract all the consumer surplus. The former indicates a collective action failure because one laser machine could serve multiple markets and the total gain in producer surplus can exceed the initial cost. These structural results rationalize the experimental findings, and also point to the importance of understanding the role of market structure in the presence of information problems.

Using the structural model, I conduct counterfactual exercises to examine the role of firm size, market competition and government policies. Since this is a second-best world with multiple imperfections, namely asymmetric information and market power, the welfare implication of increased competition is theoretically ambiguous because the two imperfections could counteract: while market power distorts quality and markup, it can also internalize the return of investing in reputation. The goal of these counterfactual exercises is to examine this interaction, and highlight some general economic forces and trade-offs faced by policy makers in regulating markets with information problems.

The counterfactual results indicate that information frictions lead to significant under-provision of quality, compared to the equilibrium outcome under symmetric information. Providing sellers (one per market) with the new branding technology could mitigate this distortion. The gain in consumer surplus, taking into account the learning process, is large, as a result of both an enlarged choice set and improved matching—high-valuation consumers self-sort into buying higher quality, though more expensive, product. While an individual seller would not undertake such costly investment because of her small market size and the difficulty in capturing the consumer surplus, a third-party could invest in the new technology and subsidize it for sellers to improve society's welfare. Alternatively, since sellers' net profits scale up with their market size, the results suggest that there could be a profitable entry opportunity for a larger upstream firm to invest in this new technology and build up a good reputation over time. Finally, to further shed light on the effects of market competition under asymmetric information, I conduct two counterfactual policy experiments wherein all sellers in one market are provided with the new technology and compete for demand and wherein the government further introduces a price regulation. The results show that the social planner would indeed want to set a higher markup to ease the competitive pressure in order to encourage effort. In other words, market competition among small firms helps to expand sales, but it can also discourage quality improvements.[1]

This study is motivated by the extensive body of economics literature on the role of advertising. Under information asymmetry, costly advertisements could act as a signal for product quality.[2] Though theoretically possible, identifying the causal effects of advertising on consumers' perception has remained an empirical challenge as it is difficult to obtain exogenous variation in the level of advertising in real-world settings. Prior studies use stated preferences elicited from lab experiments, or estimate structural models using non-experimental purchasing data. In this study, I combine the experimental variation with a structural model of learning to recover consumers' perception of product quality under different branding technologies using actual purchasing data.[3]

Studies examining the effects of branding also constitute a very large body of literature in marketing science (see review papers by Keller and Lehmann (2006) and Zhang et al. (2015)). The random assignment of branding technologies in this study generates among the first experimental evidence on the effects of branding. Findings here may be of particular interest to firms in developing countries

[1] Kranton (2003) provides a theoretical framework for analyzing the effect of market competition on the incentive to produce high quality.

[2] Bagwell (2007) provides a comprehensive review of the theoretical and empirical literature on the role of advertising. Most theoretical work focuses on equilibrium predictions between advertising and quality, where quality is exogenous (for example, see Nelson (1970); Schmalensee (1978); Milgrom and Roberts (1986); Kihlstrom and Riordan (1984); Hertzendorf (1993); Horstmann and MacDonald (1994); Zhao (2000); Linnemer (2002)). This study examines sellers' endogenous quality choice and highlights a potential channel in which costly advertising could motivate quality.

[3] There is a growing literature in marketing science that uses field experiments to study this topic, with most papers focusing on internet advertising (for example, Lewis and Reiley (2011)). See Simester (2015) for a comprehensive survey.

where brand protection is weak and existing branding technologies are "contaminated" by rampant counterfeiting activities.[4] This is closely related to the theoretical underpinning on the role of *rebranding* as disrupting the negative link between consumers and the origin brand (Prasad and Dev, 2000).[5]

To estimate the demand model, this paper builds on the literature of estimating consumer learning models for experience goods. Ching, Erdem, and Keane (2013) provides an excellent review of this literature.[6] Identification is based on reported consumption experiences and purchasing decisions observed in the household panel data.[7] In particular, by using household data collected on the day that the new premium option was first introduced, I am able to estimate consumers' prior beliefs from subsequent purchasing patterns. The framework can be used for analyzing the introduction of new goods in other settings where researchers could combine market-level price data with individual-level purchasing data.

It is also important to emphasize that with the exception of Huffman, Rousu, Shogren, and Tegene (2007), empirical papers on learning for experience goods have tended to focus on learning along the horizontal dimension of taste, as in the case of pharmaceutical markets, where uncertainty lies in the match between the drugs and the patients (for example, Crawford and Shum (2005); Dickstein (2014)).[8] This paper instead examines consumer learning on the vertical dimension of quality. I further integrate the learning model of demand with a supply-side model to study the firms' incentive and endogenize the quality choice.

This paper also adds to a growing body of empirical work on seller reputation. Despite a large body of theoretical literature on this topic (Mailath and Samuelson (2013)), there are relatively few empirical studies that examine the extent to which reputational concerns discipline sellers' behavior in real-world settings. Bar-Isaac and Tadelis (2008) provides an excellent review of some recent empirical work. While several interesting papers seek to explore reputation dynamics in online trading environments (for example, see Dellarocas (2006); Jin and Kato (2006); Cabral and Hortacsu (2010)), empirical work in the offline world is relatively sparse (McMillan and Woodruff, 1999; Banerjee and Duflo, 2000; Hubbard, 2002; Jin and Leslie, 2009; Macchiavello, 2010; List, 2006; Björkman-Nyqvist, Svensson, and

[4] The issue of counterfeiting is notorious in developing countries. See for example studies by Björkman-Nyqvist, Svensson, and Yanagizawa-Drott (2013) and Qian (2008).

[5] In line with the economic literature on the role of advertising, the marketing literature has documented that branding could help to differentiate a product from other existing products (Park, Jaworski, and Maclnnis, 1986) and signal a product's quality (Maheswaran, Mackie, and Chaiken, 1992). This literature has been primarily theoretical. The empirical questions are largely unanswered—does branding actually affect consumers' perception and does it improve firms' performance outcomes? A recent empirical study by Yi-Lin Tsai and Chintagunta (2015) examines these questions in the context of the US lodging industry.

[6] A partial list includes Roberts and Urban (1988); Erdem and Keane (1996); Ackerberg (2003); Coscelli and Shum (2004); Erdem, Keane, and Sun (2008); Israel (2005); Osborne (2005); Ching (2010a); Dickstein (2014).

[7] Other recent studies have also tried to gain identification by combining choice data with direct measures of information signals. For example, see Chintagunta, Jiang, and Jin (2009); Ching and Ishihara (2010); Ching, Clark, Horstmann, and Lim (2011) and Kalra, Li, and Zhang (2011).

[8] Bergemann and Välimäki (2006) studied the optimal pricing of a newly introduced experience good with fully forward-looking buyers and sellers. In their model, uncertainties originate in each consumer's valuation of the product whereas the aggregate distribution of preferences in the population is common knowledge. The information problem, therefore, differs from the moral hazard problem studied here.

Yanagizawa-Drott, 2013). Since the bulk of goods transactions in developing countries still take place in a non-virtual setting, it is important to understand the functioning of the reputation mechanism in this realm. 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 the buyers observe.[9] This study takes advantage of a field experiment and collects data that directly measure both the sellers' behavior and the information available to consumers over time. In line with the previous studies, the findings here demonstrate that the way consumers gather information and learn affects seller's reputational incentives. However, in contrast to the settings considered in previous studies, the results show that even in a setting where purchases are frequent and information acquisition transpires quickly, a seller's reputational incentives could remain hampered by pessimistic prior beliefs. This recalls the finding in Björkman-Nyqvist, Svensson, and Yanagizawa-Drott (2013) that learning about anti-malaria drug quality in Uganda is hampered by consumers' biomedical misconceptions about malaria and such misconceptions affect the incentive to sell fake drugs. Although the contexts differ, the policy conclusions are remarkably alike: to motivate a high quality provision, policies that could enhance consumer learning or entry of large firms may be needed.[10]

The remainder of this paper is organized as follows. Section 2 describes the setting. Section 3 outlines a conceptual framework. Section 4 describes the experimental design and the data. Section 5 presents the experimental results, which motivate the empirical model in Section 6. Section 7 uses the structural estimates to examine the welfare implications of information frictions, fragmented markets and government policies. Section 8 concludes. Additional details on data collection and technical details are available in online appendices.[11]

2 The Local Watermelon Markets in China

I begin by describing four stylized facts about local watermelon markets in Shijiazhuang, China. These facts are supported by data collected at various stages of the fieldwork, which I describe in detail below.

Fact #1. The markets are highly localized with frequent repeated interactions between local sellers and consumers.

Most watermelon transactions take place in local markets—areas where sellers gather to sell many different types of daily food products. These markets are typically located near clusters of gated

[9] A few studies have exploited quasi-experimental variation in national policy changes as shocks to the informational environment and examine the effects on market equilibrium outcomes. For example, see Jin and Leslie (2009).

[10] The study is also related to the literature on under-adoption of technology found in many developing countries. Most studies have focused on aversion to experimentation as a reason for underinvesting in profitable business opportunities and technologies (Hausmann and Rodrik, 2003; Munshi, 2004; Foster and Rosenzweig, 2010; Duflo, Kremer, and Robinson, 2011; Fischer, 2013; Dupas, 2014; Bryan, Chowdhury, and Mobarak, 2014). Cost of one-time experimentation is conceivably low in this setting. The results suggest that the main obstacle to quality provision is the information problem rather than under-experimentation. Fragmented markets could prevent the take-up of a new technology that involves high fixed cost.

[11] Link to the online appendix: http://economics.mit.edu/grad/jieb/research

residential communities. Appendix Figure 1 illustrates a typical local market. On average, households in summer purchase 1 to 2 watermelons per week (see baseline summary statistics in Table 1), mostly from the local markets. In the household data collected for this study, 78% of watermelon purchases are made from the local markets and 10% are made from nearby supermarkets.[12] Given the high frequency of repeated interactions among local sellers and consumers, we would expect that there is at least room for reputation building in this setting.

While consumers do differ in horizontal taste–some buy more often from certain local sellers than other, switchings are also quite common: in the household panel purchasing data collected for this study, only 1 out of 675 households had all of the reported fruit purchases from a single designated seller. The degree of horizontal taste differentiation and the stickiness in household purchasing behavior both matter for sellers' reputational dynamics. Section 6 addresses these in more detail and discusses identification of the structural parameters from the observed switching patterns in the data.

Fact #2. Quality varies considerably across watermelons of the same breed within the same batch. While consumers find it difficult to detect the underlying quality at the point of transaction, sellers can assess quality based on less obvious observables.

This fact is supported by ample anecdotal evidence. To formally establish the presence of information asymmetry in this setting, I conducted a sorting test with 30 fruit sellers in 30 different local markets in the city. Each of them was asked to sort 10 watermelons into two piles: one for high quality and one for low quality.[13] The watermelons were randomly picked by surveyors from the sellers' stores with no obvious distinguishable differences in outlook. The same test was repeated with 5 randomly chosen local consumers in each market. Finally, quality was measured using a sweetness meter.[14] The lightest gray line of Figure 1 plots the cumulative sweetness distribution of all 300 watermelons.[15] A one-way analysis of variance shows that 70% of the variation is explained within sellers; in other words, quality varies within single batches of watermelons at each given store. The darker grey lines compare the sweetness distribution of the premium piles sorted by sellers and consumers. The CDF graphs indicate

[12] Overall, the quality of watermelons and other perishable fruits are lower in the supermarkets than in the local markets. One reason is that supermarkets tend not to replenish inventories on a daily basis. One may imagine that it can be easier for large supermarkets to establish a reputation for quality watermelons and they can directly contract with upstream farmers, forming a vertical relationship. Having said that, we can think of various contractual frictions, in particular, the monitoring costs may be very high as compared to individual vendors or middlemen.

[13] Specifying a fixed number of watermelons for each pile may wash out differences between skilled and unskilled subjects, while not doing so could lead to trivial sorting. In practice, the maximum and minimum for each pile are set to be 7 and 3 respectively. On average, sellers sorted 4.4 watermelons into the premium pile and consumers sorted 3.5.

[14] While quality for watermelons is multi-dimensional, sweetness strongly correlates with consumer's taste. In a blind tasting test in which 210 consumers were asked to compare two watermelons of high and low sweetness measures, 97% preferred the sweeter one. This also suggests that true quality is easy to assess when consumers experience the product, implying a fast arrival of quality signals.

[15] To give a sense of the scale, a sweetness difference of 0.5 matters significantly for taste. Using self-reported satisfaction ratings in the household data and data from the quality checks, sweetness above 10.5 roughly maps to a subjective assessment of "very good" (see Appendix Figure 7).

that sellers are better than consumers at assessing quality.[16] These results demonstrate that watermelon is an experience good with asymmetric information between the two sides of the markets.

Fact #3. Consumers are heterogeneous in their willingness to pay for quality.

To elicit willingness to pay for quality, households were asked in the baseline survey to consider a hypothetical situation wherein two piles of watermelons are sold in the local markets: one pile of ordinary quality sells at 1.5 RMB/Jin[17]; 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.[18] Figure 2 plots the cumulative distributions of the reported willingness to pay for households in different income and age groups. Willingness to pay is higher for households with higher income (left figure) and for non-elderly households (right figure).

Fact #4. In contrast to many other fruits sold in these markets, there is a stark absence of quality differentiation for watermelons at baseline.

Although this study focuses on downstream local markets, the lack of quality differentiation is seen at every stage of the watermelon value chain, from the farmers to the middlemen, and in the wholesale markets. Despite the underlying variation in quality within each batch of watermelons, sellers sell an undifferentiated pile and do not price watermelons by quality. Within each local market, there is also little price variation across sellers. This contrasts sharply with other fruits, including peaches, cherries, bananas, and grapes, which are also sold in these markets, and for which there are substantial quality differentiations. For example, sellers usually sell multiple bins of peaches (of the same breed) at different prices. This is true both from anecdotal observations and in the sales data collected for this study. On average, 64% of the sellers sort peaches at sale on any given day and a given seller sorts peaches for 67% of the time during the study period.

These four facts together lead to the central research question: given the considerable amount of quality variation and sellers' ability to assess the underlying quality, why is there a lack of quality differentiation for watermelons at baseline? One feature that distinguishes watermelons from many other fruits is the fact that consumers can only determine the quality after purchasing decisions have been made.[19] In contrast, the quality of other fruits can be easily observed by both sides of the markets at the point of transaction—for example, a nice peach looks different from a rotten one. This leads us to think that asymmetric information might be playing a role for the absence of quality differentiation for watermelons.

It is worth mentioning that various forms of implicit discrimination could operate in these markets.

[16]The two-sample Kolmogorov-Smirnov test strongly rejects the equality of distributions at 1% level.

[17]1 Jin $\approx$ 1.1 pounds. The rest of the paper uses Jin as the unit for price.

[18]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.

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

For example, sellers could pick higher quality watermelons and give them to customers whom they know well or repeat customers. To the extent that such preferential treatments may not perfectly align with willingness to pay, there would still be welfare losses due to the mis-matches. And to push the argument one step further, ex-ante it is possible that relational contracting has fully solved the information problem and perfectly allocated high quality watermelons to high valuation customers. If so, we would not expect to see an effect on sales outcomes when sellers were induced to sort by quality under the experiment. Section 5.3 first presents reduced form results that profits are significant higher under differentiation than that under no differentiation. After that, Section 6 goes on to estimate a structural model to gauge the distortion and welfare loss due to the lack of differentiation at baseline.

3 Model: Quality Provision with Asymmetric Information

This section sets up a conceptual framework for analyzing quality provision with asymmetric information. The framework is adapted from Shapiro (1982). I first set up the model and specify the assumptions. After that I discuss two broad explanations for the lack of quality differentiation at baseline in light of the model. Finally, I consider the role of consumers' prior beliefs and relate that to the experimental design.

3.1 Basic Setup

The supply side: A single long-run seller faces a fixed pool of consumers. Time horizon is discrete and infinite. The seller maximizes expected discounted sum of profits with discount factor $\delta \in (0, 1)$.[20]

In each period, the seller could choose to sell just one "normal" product, or he could choose to introduce a new "premium" product and sell both. I call the former "no differentiation" and the latter "differentiation."

The per-unit cost (P_W) and price (P_N) of the normal product are assumed to be fixed. Let $\underline{\gamma}$ denote the quality of the normal product, where quality is operationalized as the probability that a consumer finds the product satisfactory. $\underline{\gamma}$ is exogenously fixed and known by consumers.

If the seller chooses to introduce a premium product, she chooses the quality γ_H, which is initially unobserved by consumers.[21] The extra marginal cost of the premium product is $C(\gamma_H; \underline{\gamma})$, where $C_{\gamma_H} = \partial C/\partial\gamma_H > 0, C_{\gamma_H\gamma_H} = \partial^2 C/\partial\gamma_H^2 > 0$. The seller also sets the price of the premium product, denoted as P_H^t. To focus on the seller's optimal policies of the premium option, I assume that the

[20]The model here abstract away from market competition. The assumption is made to match the experimental setting: in each market, only one seller was incentivized to differentiate quality, and there was little strategic response from the others (see Section 6). Therefore, I work with a monopoly model here and defer the counterfactual analysis of oligopolistic competition to Section 7.

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

price and quality for the normal product are held the same as that under no differentiation. The main qualitative takeaways from the model do not hinge on this assumption. In the empirical analysis, I shall take a closer look at sellers' actual quality and pricing behavior when they start to differentiate quality as incentivized by the experiment.

The demand side: There are many ways that one could model consumers' behavior and beliefs when the seller introduces a premium option. The model here focuses on the aspect of consumer learning, which we may expect to play an important role for newly introduced experience goods. Section 5 presents some empirical evidence that is consistent with consumer learning. In this setting, since consumers are not informed about the experiment, it is therefore plausible from their perspective to regard the new product as having some fixed underlying quality, which is initially unknown but can be learned over time via actual consumption experiences.

To model the learning process, I adopt a similar framework to that in Dickstein (2014). In each period, consumers share some common beliefs regarding the quality of the premium product. Purchase decisions are made based on the current beliefs, which are updated after each period's consumption experience. Suppose that the prior beliefs about γ_H follow a beta distribution with parameters (a_0, b_0). Such prior beliefs can be formed from various past experiences that consumers have had with related products (see discussions in Section 4.2). The seller's initial reputation can be proxied by the prior mean, which is given by $\mu^0 = \frac{a_0}{a_0+b_0}$. Let e_t denote period t's experience realization, which is a Bernoulli random variable with success (satisfactory) probability γ_H. For analytical tractability, I assume that all consumers receive the same experience shock e_t in each period when they purchase the premium product and that information is shared to those who do not purchase by word of mouth.[22] Since beta distribution is the conjugate prior for Bernoulli likelihood, beliefs in period t, after a sequence of experience realizations $\mathbb{e}^{t-1} = (e_1, \ldots, e_{t-1})$, simply follow a beta distribution with parameters $(a_0 + s_{t-1}, b_0 + f_{t-1})$, where s_{t-1} and f_{t-1} are the number of satisfactory and non-satisfactory experiences up to time $t-1$.[23]

In each period, consumers either buy one unit of the product or do not buy any product at all. The utility of not buying is 0. Consumers' valuation is uniformly distributed between $[\underline{\theta}, \overline{\theta}]$ with mass M. For a consumer with valuation θ who buys a product at price P, the utility is $\theta - P$ if the product is satisfactory and $-P$ if it is not. Consumers are assumed to be risk neutral, and in each period, they

[22]In reality, consumers receive different experience shocks in each period and it is more natural to think of γ_H as the mix of good watermelons at a given point in time. In the structural estimation, I enrich the model by allowing individuals' beliefs to diverge over time with observed experience realizations in the data.

[23]The model assumes a naive Bayesian updating process. One enlightened consumer could do better if she could figure out the seller's actual choice of γ_H; if all consumers are enlightened, the model corresponds to the pure moral hazard model in Klein and Leffler (1981). I present and discuss this alternative model in Appendix B. The alternative model admits a continuum of trigger-strategy equilibria (with different implicitly contracted quality). It offers an alternative explanation for the lack of quality differentiation at baseline, i.e. a coordination failure between sellers and consumers. Having said that, the alternative framework does not speak about equilibrium switching dynamics, thus it would be hard to explain the learning patterns observed in the data.

make their purchase decisions to maximize the current period's expected utility. The model abstracts away from the option value of experimentation in order to match with the structural model. I discuss the challenges for modeling forward-looking behavior in Section 6.

The seller's problem: The seller chooses whether to differentiate by quality or not.[24] Let $Q^t_{N,\text{nodiff}}$ denote the demand under no differentiation, $Q^t_{H,\text{diff}}$ and $Q^t_{N,\text{diff}}$ denote the demand for the premium and normal products under differentiation. Under no differentiation, the seller's discounted sum of profits are fixed, given by the parameters of the model:

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

Under differentiation, the seller faces a dynamic demand system. In particular, $Q^t_{H,\text{diff}}$ and $Q^t_{N,\text{diff}}$ are functions of $\mu^{t-1}(\mathbb{e}^{t-1}(\gamma_H); a_0, b_0)$, which evolves over time as consumers learn.[25] For a given γ_H, the optimal P^t_H is imposed by static profit maximization. (Because the stylized model assumes complete information diffusion, there is no dynamic implication of current sales. See discussions below). The expected discounted sum of profits under γ_H is

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

where the expectation is taken over sequences of experience shocks $\{e_t\}^{\infty}_{t=1}$ generated by γ_H. Let γ^*_H denote the argmax of $\Pi_{\text{diff}}(\gamma_H)$ and $\Pi_{\text{diff}}(\gamma^*_H)$ the maximized expected value under differentiation.

Suppose there is an initial fixed cost F of introducing a premium option, and the seller chooses to differentiate if and only if $\Pi_{\text{diff}}(\gamma^*_H) - F > \Pi_{\text{nodiff}}$.[26]

This completes the setup of the model. In Section 6, I provide some descriptive evidence on the model's key assumptions and structurally estimate the model. For the structural estimation, I enrich the basic setup by incorporating greater dimensions of consumer heterogeneity, private experience shocks, information diffusion, and market competition. For the remainder of this section, I work with the basic

[24] Theoretically, it is possible that the seller's profit maximization decision is to only sell the premium product. This could happen if the cost of providing quality is very low. By introducing an inferior option, the seller will lose some sales on the premium product. However, such behavior is not observed in the data. Thus I exclude this case for convenience.

[25] Demands are determined by cutoff types as in standard vertical taste models. For interior solutions, we have:

$$\begin{aligned}
Q^t_{H,\text{diff}}(P^t_H, P_N, \mu^{t-1}(\mathbb{e}^{t-1}(\gamma_H); a_0, b_0); \underline{\gamma}) &= \left(\overline{\theta} - \frac{P^t_H - P_N}{\mu^{t-1}(\mathbb{e}^{t-1}(\gamma_H); a_0, b_0) - \underline{\gamma}}\right)\frac{M}{\overline{\theta} - \underline{\theta}} \\
Q^t_{N,\text{diff}}(P^t_H, P_N, \mu^{t-1}(\mathbb{e}^{t-1}(\gamma_H); a_0, b_0); \underline{\gamma}) &= \left(\frac{P^t_H - P_N}{\mu^{t-1}(\mathbb{e}^{t-1}(\gamma_H); a_0, b_0) - \underline{\gamma}} - \frac{P_N}{\underline{\gamma}}\right)\frac{M}{\overline{\theta} - \underline{\theta}}
\end{aligned}$$

where a_0 and b_0 are given.

[26] F is not needed for deriving the comparative statics. Without that, if non-differentiation is the optimal strategy for the monopolist under asymmetric information, it is also the optimal strategy under symmetric information as well as under the first best because only the highest valuation ($\overline{\theta}$) matters for the decision on the extensive margin. However, this is a knife-edge scenario. In reality, any positive initial costs of introducing the premium product could break this.

framework to derive some testable implications, which motivate the experimental design and guide the reduced form analysis.

3.2 The Effects of Prior Beliefs

In light of the model, there are two broad explanations for the lack of quality differentiation at baseline: high costs and asymmetric information. First, if cost is high relative to consumers' valuation for quality, $\Pi_{\text{diff}}(\gamma_H^*) - F$ could be smaller than Π_{nodiff}. Suppose $\overline{\theta} < C(\underline{\gamma};\underline{\gamma}) + P_W$, i.e., the highest consumer valuation is lower than the marginal cost of providing higher than the minimum quality $\underline{\gamma}$, then higher quality will not be demanded and supplied even under symmetric information. Second, under asymmetric information, a good reputation may take a a long time to establish if prior beliefs are pessimistic.[27] Therefore, the seller who rationally discounts the benefits of a good reputation in the future may lack the incentive to invest in reputation building. Hence, the market can be stuck in an outcome with no quality differentiation.

In reality, these two aspects will act jointly. Better prior beliefs enhance the seller's return of building reputation (Proposition 1 below), but how good that needs to be to induce quality differentiation depends on costs (relative to consumers' valuation). Having said that, the welfare implications of the two stories are very different: for the former, the distortion on quality provision caused by the information problem is small, whereas for the latter it could be large.[28]

In practice, it is hard to directly infer costs. Therefore, to understand the main barrier for quality provision, the experiment seeks to create exogenously variations in prior beliefs. These variations should have minimal effects if the key barrier for quality provision is high cost. On the other hand, if the information problem is the key barrier, changing prior beliefs could significantly affect sellers' reputational incentive and hence their quality differentiation behavior. The effects are stated in the following two propositions:

Proposition 1: *(Incentive to provide quality)* For a given discount factor δ, a distribution of consumers' valuation θ and a marginal cost function $C(\gamma_H;\underline{\gamma})$, $\Pi_{\text{diff}}(\gamma_H)$ increases with a_0 and decreases with b_0.

Proposition 1 says that improving prior beliefs, either by increasing a_0 or decreasing b_0, raises the seller's return under differentiation. The intuition is straightforward because good initial reputation is always a positive asset for the seller. In particular, holding a_0 fixed, a lower b_0 implies a higher prior mean and a larger prior variance, and hence a faster speed to establish trust and larger discounted

[27] The model assumes that off-equilibrium beliefs are pessimistic. The theoretical literature does not provide clear guidance on beliefs off the equilibrium path in extensive form games. For instance, Fudenberg, Kreps, et al. (1988) and Fudenberg and Levine (1993) argue that it may only be reasonable to restrict beliefs to be correct on the equilibrium path.

[28] In general, the welfare implication of information frictions is theoretically ambiguous in a setting with market power. If the monopoly's profit-maximizing quality under symmetric information exceeds the socially efficient quality level, the lack of information might help by moving the monopoly's choice closer to the efficient level. In Section 7, I use the structural estimates to examine the counterfactual quality level with symmetric information and with and without market power, and I use that to quantity the distortion due to the two imperfections.

returns. The next proposition examines how the seller's optimal quality choice responds to prior beliefs if she differentiates.

Proposition 2: *(Optimal quality choice)* For a given discount factor δ, a distribution of consumers' valuation θ and a marginal cost function $C(\gamma_H; \underline{\gamma})$, if $\frac{\partial^2 \Pi_{\text{diff}}(\gamma_H^*; a_0, b_0)}{\partial \gamma_H \partial a_0} > 0$, γ_H^* increases with a_0; similarly, if $\frac{\partial^2 \Pi_{\text{diff}}(\gamma_H^*; a_0, b_0)}{\partial \gamma_H \partial b_0} < 0$, γ_H^* decreases with b_0.

Proposition 2 states a simple monotone comparative statics: if prior beliefs and quality are complementary, the seller will be induced to provide higher quality when prior beliefs improve.[29] In reality, this condition should be much less stringent with sales effects: optimistic beliefs encourage sales, which enables information to spread faster, and thus rewards good behavior and punishes bad behavior faster. This channel is absent here because the basic setup assumes perfect information diffusion. In the structural estimation, I shall take into account private experience shocks and investigate the role of information diffusion.

Section 4 describes the experiment and relate the treatments to this framework for thinking about their effects. After that, Section 5 tests the two predictions of the model.

4 Experimental Design and Data Collection

4.1 Experimental Design and Timeline

Overview. The experiment was conducted in Shijiazhuang, the capital city of Hebei province, China.[30] The city has over 800 gated communities and more than 200 local markets. Randomization happened at the local market level. 60 sellers located in 60 different markets were recruited to participate in the study following an initial screening and a sequential selection procedure to minimize heterogeneity in the study sample for power concerns and logistical purposes. Details for the screening process and selection criteria are described in Appendix C.1.

There are typically 3 to 5 sellers in each local market, but only 1 was selected. In what follows, I call the 60 participating sellers the *sample sellers*, as opposed to the *other sellers*, who operated businesses in these markets but who were not directly involved in the experiment. All sample sellers were asked to experiment with selling two piles of watermelons: a premium pile and a normal pile. Sellers were free to set quality, price, and quantity for each pile. They were randomized into 6 groups.

Branding treatments. Sellers were randomized into one of the three branding groups: laser, sticker and label-less. Every morning, surveyors visited the sellers' stores and performed a free branding service. For the laser group, surveyors used a laser-engraving machine to print a laser-cut label of the words "premium watermelon" ("Jing Pin Xi Gua" in Chinese Pinyin) on the watermelons in the premium pile.

[29] In general, γ_H^* is non-monotonic in a_0 and b_0. When $a_0 + b_0$ is sufficiently large, as one of the two parameters tends to 0 (i.e. very pessimistic or very optimistic beliefs), the seller's incentive to build a reputation vanishes (i.e. γ_H^* goes to $\underline{\gamma}$).

[30] Urban area: 399.3 sq km (154.2 sq mi); urban population: 2,861,784; urban density: 7,200/sq km (19,000/sq mi)

For the sticker group, surveyors pasted a sticker with the same words. For the label-less group, surveyors did nothing. It is important to emphasize that the branding treatment was only for watermelons in the premium pile; those in the normal pile were left as they were. Figure 3 shows pictures of the branding treatments. Most sellers sold two piles of watermelons at the beginning of the experiment, but some reverted back to non-differentiation after some time. For those sellers, branding was withdrawn because there was no longer a premium pile.

A cross-randomized incentive treatment. Within each branding treatment group, half of the sellers were randomly given an incentive to maintain quality for the premium pile. The incentive treatment was enforced via unannounced quality checks twice per week. At every check, surveyors randomly picked one watermelon from the premium pile and one from the normal pile. The quality of both was measured using a sweetness meter (see Appendix Figure 2). For sellers in the incentive group, if the sweetness of the former attained 10.5 at both checks, sellers received a monetary reward of 100 RMB at the end of the week. (A seller's average daily sales profit is around 100-200 RMB.) Sellers in the non-incentive group received the same quality checks, but were not given any reward. The incentive was removed in the later part of the intervention, and that was unanticipated by the sellers.

In total, there were 6 distinct treatment units: three branding treatments crossed with an incentive treatment. Randomization was stratified on housing prices, i.e. a dummy variable indicating whether the baseline average housing price in the surrounding gated communities is below or above the median. Appendix Figure 3 shows a map of the 60 sellers, marked by groups. Note that these markets are geographically segregated and the average distance between any two markets is 3 kilometers. Since watermelon transactions are highly localized, spillover effects across markets should be minimal.[31]

Figure 4 describes the timeline of the study. The market intervention was rolled in from July 13 to July 19, 2014. There are numerous issues with sellers' sales recording at the beginning. Thus, for the subsequent analysis, I exclude data for the rolling-in period and define July 19 to be day 1 of the intervention. At the start of the intervention, most sellers, except for three, sold two piles of watermelons. A few reverted back to non-differentiation after several days. On August 3, two weeks into the intervention, a universal announcement was made to all sellers that they were free to decide whether they wanted to continue with quality differentiation or not. On August 23, six weeks into the intervention, the incentive was removed. Finally, the intervention was phased out from September 6 to September 12. An endline survey was conducted upon the surveyors' final visit to sellers' stores. Two follow-up surveys were conducted after the market intervention ended to examine longer-term outcomes.

[31] In principle, there could still be spillovers even if demands are localized. For example, sellers may communicate with one another and adopt similar strategies. Using the social network information collected from the 60 sellers in the study sample, I find little correlation between the price charged for watermelons in the premium pile by a given seller and the average of the price charged by their friends (conditional on treatment). There is a positive correlation among geographic neighbors, most likely reflecting similar costs.

4.2 Testable Implications

To predict the effects of the treatments on sellers' behavior and market outcomes, I relate the experimental design to the framework in Section 3 and highlight several key features of the different branding technologies in the Chinese context for thinking about their potential effects on consumers' prior beliefs.

Two aspects of laser branding are distinct from sticker branding and the label-less case. First, the cost of laser machines is very high (around 50 to 60k RMB, or 8 to 10k USD), and from anecdotal evidence, consumers seem to share a common understanding that laser branding is more expensive than sticker branding. In this aspect, laser branding represents a large conspicuous sunk investment. Klein and Leffler (1981) discusses the role of the initial sunk investment in a situation where consumers are uncertain about underlying costs. The argument is known as forward induction: upon seeing a costly sunk investment, consumers think that such an investment can only be profitable if the seller's future quasi-rents are large; however, such quasi-rents would be lost if the seller shirks quality today, and therefore she must exert effort. In other words, conspicuous advertising investments signal the presence of a price premium that is high enough to motivate high quality. The argument therefore suggests that consumers' prior beliefs could be more optimistic under laser.

The forward induction argument is an equilibrium refinement argument (formalized in Appendix B). In reality, it may be hard for sellers and consumers to immediately coordinate on a new equilibrium with the introduction of laser-branded watermelons. A more realistic model would incorporate consumer learning to shed light on market dynamics. I highlight a second aspect of laser branding that is distinct from sticker branding in the context of China: the former is a completely novel branding technology whereas the latter is highly "contaminated" due to rampant counterfeiting activities in the past. This is also true in many other developing countries where brand protection is weak. Evidence from the consumer pilot survey suggests that consumers regard laser branding as being more effective in deterring counterfeits than stickers, which can be easily fabricated. Therefore, laser branding could potentially dispel the negative historical stereotypes associated with stickers. This discussion is similar to the collective reputation model studied in Tirole (1996). A key takeaway from the model is that equilibrium could be history dependent. After many episodes of past bad behavior, the group could be stuck in a bad-reputation equilibrium because the low collective reputation makes good behavior today a low-yield individual investment, which in turn breeds low collective reputation in the future. A new "uncontaminated" branding technology could potentially wipe out bad stereotypes from history, thereby allowing trust to establish faster. Proposition 1 and 2 then suggest that sellers in the laser group may have a stronger incentive to build reputation and provide higher quality.

Another potential effect of laser branding is that it may represent something "cool", and therefore has a direct effect on brand prestige other than affecting consumers' prior perception about quality. Section 6.3 investigates this by incorporating and estimating a product-specific constant ν for the premium option under laser label in consumers' indirect utility function to account for this possibility.

Finally, we can think of the incentive treatment as shifting the posterior beliefs at the end of the incentive period. The idea is based on the following: if the incentive could motivate sellers to provide higher quality, then upon its removal after some period T, sellers who have had the incentive are essentially endowed with a higher reputation than those sellers who have not had the incentive. Proposition 2 suggests that higher quality may be sustained even in the post-incentive period. This, therefore, provides a direct test for whether the reputational forces at work are consistent with the model's prediction.

4.3 Data Collection and Description

Baseline surveys. The seller and household baseline surveys were conducted in July 2014. Table 1 summarizes the sample characteristics. On average, a local market houses 3 to 5 fruit sellers. Most sellers sell fruits all year long and do not expect to relocate. The median household consumes 1 watermelon per week in the summer, and 75.6% of the households list the local market as the main source for watermelon purchases. These patterns support the motivating facts described in Section 2.

Quality. Random quality checks were carried out twice a week as described in Section 4.1. For each watermelon, sweetness was measured both at the center and at the side. For the empirical analysis, I take the average sweetness as the proxy for a watermelon's quality.

Pricing. Surveyors visited the local markets everyday and recorded the quality differentiation and pricing behavior of all sellers in these markets as well as the daily wholesale price.

Sales. Sellers were asked to record their daily sales information for watermelons and peaches. For each transaction, sellers were asked to record the fruit type (watermelon or peach), sales quantity (in Jin), sales values (in RMB), and the corresponding quality category, premium or normal, if the sold fruit is watermelon.[32] Omissions and errors in recording were unavoidable, and occasionally sellers had to lump several sales together if they happened around the same time. It would be of concern if for some reason the noise in recording differs systematically across the treatment groups. To check this possibility, a second source of sales information was collected starting from mid-August. In particular, besides the transaction-level records on each day, sellers were also asked to recall the total sales quantity of the previous day. As a first pass, the difference between the self-recalled and the recorded total sales quantity does not differ significantly across the treatment groups.[33]

[32] On the recording sheet, sellers were also asked to distinguish between different breeds of watermelons. For all of the subsequent analyses, I focus on the most popular breed, called "Jingxin" in Chinese. Sales of all the other breeds constituted less than 2% of the total recorded sales.

[33] A related concern is that there might be differential recording noises by quality categories across the treatment groups even though the aggregate sales of the two piles do not differ. To examine this concern, I compare the daily sales quantity of the premium pile recorded by sellers with that inferred from the surveyors' records. In particular, on each day before surveyors carried out the branding service, they counted the number of branded watermelons left from the previous day and the number of newly branded ones. Using this information, I could back out the number of branded watermelons sold on a given day. While the timing difference between the branding service and the collection of the recording sheet

In total, there were 60,806 transaction records in the period from July 19 to September 6, 2014. 81% of transactions were for watermelons and 19% were for peaches. The mean quantity (measured in Jin) per watermelon transaction is 15.6, approximately the weight of a typical watermelon. On average, sellers sell 257 Jin of watermelons every day, and the average daily sales profit is 103 RMB. Here and in all subsequent analysis, sales profit is computed using sales quantity and prices.[34]

Household purchasing. 675 households in 27 communities, evenly distributed across the treatment groups, were recruited to record the family's summer fruit consumption experiences. The recruitment process is explained in detail in Appendix C.1. For each fruit purchase and consumption experience, households were asked to record the date of the purchase, the place of the purchase, the quantity bought, the amount paid, and a satisfaction rating ranging from 1 to 5, where a higher number indicates higher level of satisfaction. Importantly, households were asked to indicate whether the purchase was made from the sample seller or from other places (including other sellers in the local market, nearby supermarkets and other places) and whether the purchased fruit had any branding on it or not.

In total, there were 15,292 purchase records from the 675 households in the study sample over 8 weeks. 30.8% of the purchase records were for watermelons, of which 78% were purchased in local markets. The median for the number of watermelons consumed per week is 1 and the mean is 1.15 with standard deviation 1.06. These numbers match well with that in the baseline survey as shown in Table 1.

An important issue with the household recording data is that information for purchase place, branding and satisfaction rating is missing for some purchase records. I infer some of the missing information by matching the household data with the seller data, and drop households with three or more missing weekly records. The primary goal is to minimize the bias associated with analyzing purchasing patterns in which missing values would be treated as non-purchase. Appendix C.2 discusses the issue and the cleaning procedures in detail. The final analysis sample consists of 4,309 watermelon purchase records from 573 households in 26 communities. Characteristics of the final analysis sample look very similar to those in the full sample (see Appendix C.2).

Endline surveys. The seller endline survey was conducted during the surveyors' final visit to the sellers' stores. Sellers were asked about their planned future quality differentiation decisions as well as their willingness to pay for different branding technologies. The household endline survey was distributed and collected together with the last week's recording sheet.[35] To examine changes in perceptions, the

introduces some additional noise, the finding that the correlation between the two measures does not appear to differ between the laser and sticker groups serves as a first pass and alleviates some of the concerns for differential recording noises across groups that may drive the empirical results.

[34]Sales profit = premium pile price × premium pile sales quantity + normal pile price × normal pile sales quantity - total sales quantity × wholesale price. Alternatively, I can use the recorded sales values to calculate profits. The latter measure contains more missing values. Results are qualitatively robust.

[35]Overall, 10% of the households did not turn in the last week's record and the endline survey. Characteristics of households with missing endline data look similar to those who turned in and do not differ across groups.

same question to elicit willingness to pay for quality was asked again, but this time for watermelons under three different branding technologies. Specifically, households were asked to compare two piles of watermelons sold at the local market, one of ordinary quality at 1 RMB/Jin and the other of premium quality (as determined by the seller) but at a higher price. Households were asked to consider three scenarios of laser branding, sticker branding and no branding (label-less) for the premium pile, and for each, they were asked to indicate the highest price they were willing to pay for the premium option.[36]

Post-experiment follow-up surveys. To examine longer-term outcomes, two follow-up surveys were conducted one week and one year after the intervention. Surveyors collected information on quality differentiation and pricing behavior of both the sample sellers and the other sellers in these markets. Attrition rate is small: 1 seller dropped out during the intervention because the market was closed for road construction. For the second follow-up, surveyors were able to locate 57 of the original 60 sellers.

4.4 Balance Checks

Appendix Tables 1 to 3 present the balance checks on community, seller and household baseline characteristics. The mean for the label-less non-incentive group is shown as the constant and the OLS regression coefficients for the other five group dummies are reported. The last column reports the p-value for the Wald test of joint significance. Overall, only 1 out of 105 coefficients is statistically significantly different at the 10 percent level, 7 at the 5 percent level, and 2 at the 1 percent level. The joint test cannot reject no significant differences across groups for 16 of the 21 variables at the 10 percent level.

5 Experimental Evidence on the Effects of Branding and Incentive

This section provides experimental evidence on the effects of the branding and the incentive treatments, and tests the predictions of the model in Section 3.

Figure 5 plots the number of sellers who differentiated quality for watermelons at sale in each treatment group over time. We see that most sellers sold two piles during the first two weeks. However, the number drops drastically for the label-less group after the announcement. On the other hand, most sellers in the sticker and laser groups continued to practice quality differentiation until the end of thre intervention.

To understand why sellers who were induced to differentiate quality had not already done so at baseline, the rest of the section examines the demand side's response and sellers' quality, pricing and sales. Section 5.5 reconciles these reduced form findings with the lack of quality differentiation at baseline and provides a qualitative explanation.

[36]The reference prices for the normal pile option were also different in the baseline and endline surveys to match the actual average market price at the time when the survey was conducted.

5.1 How Do Different Branding Technologies Affect Consumers' Prior Beliefs?

This question is difficult to examine in a regression framework. A household's purchase decision in a given period is affected by the entire history of purchasing and consumption experiences, which are in turn affected by prior beliefs. In Section 6, I use a structural approach to formally model the prior distribution and the learning process to shed light on this question. In this section, I provide some reduced form evidence that is consistent with the model's predictions as discussed in Section 4.2.

In so doing, I collapse the household panel data to household-week level and create a dummy variable for whether the household purchased any premium watermelons in a given week. I regress the purchase dummy on two measures of past experiences: (1) the average lagged satisfaction rating of all premium watermelons purchased prior to the given week and (2) the percentage of past experiences that attain the highest satisfaction rating. Note that if a household has never purchased any premium watermelons in the past, these lagged experience measures are not defined. Therefore, the coefficients are only estimated from household-week observations conditioning on a positive number of premium watermelons being purchased prior to that given week.

Results are shown in Table 2. For this analysis, I focus on the sticker and laser groups because sellers in the label-less group quickly reverted back to non-differentiation and there was no longer a premium pile. All regressions control for time fixed effects as well as household baseline characteristics.[37] Only the coefficients for the lagged experience measures are reported. Column 1 and 2 of Panel A show that lagged experiences strongly predict repurchasing decisions for households in the laser markets. To interpret the magnitudes, take the estimate in column 2, which shows that for two similar households at a given point in time, the household that has had only very good past experiences is 45% more likely to repurchase a premium watermelon than the household that has not had any very good experiences (but that has experienced the product). On the other hand, the coefficients are much smaller and noisier for households in the sticker groups, as shown in columns 3 and 4. These patterns are consistent with the discussions in Section 4.2: prior beliefs may be more "stubborn" under stickers, which implies that purchasing decisions would be less responsive to past consumption experiences.

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

5.2 How Do Different Branding Technologies Affect Sellers' Quality Choice?

The descriptive patterns in Table 2 suggest that it is easier to establish trust under laser than under sticker. Proposition 2 says that faster learning could enhance sellers' reputational incentive and motivate higher quality provision. To test this prediction, Panel A of Table 3 compares the average premium pile

[37] These characteristics include the household size, % of elderly, monthly income, the baseline self-reported average number of watermelons consumed per week and the willingness to pay for quality (in RMB/Jin).

quality, measured in sweetness, for sellers in the sticker and laser groups. I pool together the quality check data for the two groups and regress the measured sweetness of the premium watermelons on the laser group dummy. Standard errors are clustered at the seller (market) level, which is the unit of randomization. This applies to all the regression analysis below unless otherwise stated. We can see that on average, sellers in the laser group provide significantly higher quality than sellers in the sticker group, and this is true both with and without the incentive.

To further understand sellers' quality differentiation behavior, I look at how the quality of the premium pile compares with that of the normal pile, and at how the two compare with the market average. Specifically, I run the following regression:

$$y_{ipt} = \alpha + \beta \text{Premium}_p + \gamma_i + \lambda_t + \epsilon_{it} \tag{3}$$

The outcome variable y_{ipt} is quality, measured in sweetness, for pile p of seller i at time t during which the seller sold two piles. Time is day for price and week for quality. The key explanatory variable is a dummy for the premium pile. Thus, α represents the mean of the normal pile and β measures the average difference between the two piles. To focus on the effect of the branding treatment, I restrict the sample to sellers in the non-incentive groups and estimate Equation 3 separately for the laser and sticker groups. The regression controls for seller (γ_i) and time (λ_t) fixed effects.

Results are presented in Table 4. Panel A shows that the average quality of the premium pile is significantly higher than that of the normal pile for both the laser and sticker groups. However, the difference could be either due to a genuinely better quality of the premium pile (relative to the market average) or a deterioration of quality for the normal pile. To examine the latter possibility, Panel B runs the same regression but with quality difference from the market average as the outcome variable. Since sellers source from the same wholesale market and thus face the same pool of watermelons on any given day, I use the average sweetness of the randomly picked watermelons from sellers in the label-less group after they had reverted back to non-differentiation (i.e., from week 3 onwards) as a proxy for the average quality.[38]

Results in column 3 shows that sellers in the laser group maintained a higher quality for the premium pile and kept the normal pile quality on par with the market average.[39] The fact that the normal pile quality is not compromised for the laser group suggests that sellers may have spent more effort on sourcing good watermelons in the upstream. This result alone shows that reputational incentive is present in this setting and could discipline sellers' behavior. As long as providing higher quality involves positive efforts, in a one-period game, sellers would not exert such additional efforts and would

[38]Every round of quality checks were conducted on different days across sellers, but the timing difference was within one or two days. As long as the quality of the pool does not fluctuate very much from day to day, the constructed average quality can be seen as a proxy for the average quality of the pool faced by a seller at a particular quality check.

[39]The p-value for a one-sided test for the normal pile being worse than the market average is 0.356.

just randomly label some watermelons as "premium" and sell them at a higher price.

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

5.3 How Do Different Branding Technologies Affect Sellers' Return?

Table 5 examines how different branding technologies affect sellers' sales outcomes. To focus on the effects of the branding treatments, I restrict the sample to the non-incentive groups and run the following regression:

$$y_{it} = \alpha + \beta_1 \text{Sticker}_i + \beta_2 \text{Laser}_i + \lambda_t + \epsilon_{it} \tag{4}$$

The outcome variables are log sales profits (in RMB), markup from market average price (in RMB/Jin)[40] and sales quantity (in Jin) for each pile, and the total sales quantity for seller i on day t. If a seller stops differentiating quality, the unit price for the premium pile is defined to be the same as that for the normal pile and sales quantity for the premium pile is coded as 0. Results are shown in Table 5. Sticker and laser are group dummies and the omitted group is the label-less group. All regressions include day fixed effects (λ_t) to control for time-specific aggregate shocks, such as weather. The even columns control in addition for community and seller baseline characteristics.

Column 1 and 2 show that on average, the laser group earns 30-40% higher sales profits than the label-less group. This is due to both a higher unit price (columns 3 and 4)[41] and higher sales quantity for the premium pile (columns 5 and 6). Sales of the normal pile are not significantly different from those of the label-less group. The results suggest that under quality differentiation, sellers in the laser group attract more high-end customers to buy the premium product without losing sales on the normal product. On the other hand, for the sticker group, sales of the premium pile appear to be lower on average than the laser group (the p-value of a one-sided test is 0.238) despite a lower markup. Furthermore, sales of the premium pile (columns 5 and 6) are offset by a reduction in the sales of the normal pile (columns 9 and 10). As a result, total sales and profits are not significantly different from

[40] Here and in all subsequent analysis with prices, I use the listed prices as observed by surveyors during the morning visits to the markets. Alternatively, for the sample sellers, I can use the effective prices, which are calculated as total daily sales values divided by total daily sales quantity for each quality category. Results look very similar and all the qualitative conclusions are robust.

[41] The average wholesale price during this period is 0.62 RMB/Jin; thus the price markup is quite significant.

those of the label-less group.[42] Overall, the results are consistent with Proposition 1: laser branding could enhance consumers' prior beliefs, thereby enhancing sellers' return under quality differentiation.

5.4 How Is Seller's Quality Choice Affected by the Incentive Treatment?

As discussed in Section 4.2, the incentive treatment provides a direct test of the model's prediction: if the incentive had successfully facilitated sellers' initial reputation building, then higher quality could be sustained even after the incentive was lifted. To test the prediction, I first provide some suggestive evidence for the former statement.

Panel B of Table 3 runs the same OLS regression as that in Panel A but with a dummy for the incentive treatment and separately for the sticker and laser groups. I restrict the time to the incentive period (i.e. the first six weeks). We see that the incentive did lead sellers to provide higher quality.

Given this, we expect that over time as consumers experience the product, the incentive groups should outperform their non-incentive counterparts, especially under laser label where learning is salient (see Table 2). To investigate the time dynamics, I fit a linear time model (see Appendix Figure 4 that plots the group average quantity over time in the raw data):

$$\text{Premium Quantity}_{it} = \alpha + \beta_1 \text{Time}_t \times \text{Incentive}_i + \beta_2 \text{Time}_t + \gamma_i + \epsilon_{it} \tag{5}$$

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

The results above imply that sellers in the laser incentive group should have established a higher reputation than their non-incentive counterparts at the time when the incentive was lifted. While data on market perceptions at each point in time are not available, households' perceptions elicited in the endline survey could be suggestive. Table 7 shows the results of regressing households' self-reported willingness to pay under different product differentiation technologies on treatment group dummies. The omitted group is the label-less non-incentive group and the even columns control in addition for household baseline characteristics. We see that the average willingness to pay is the highest under the laser label, and it is the highest for households in the laser incentive markets. Households in the sticker incentive markets also appear to be willing to pay more under the sticker label, but the estimate is noisy and the magnitude is much smaller.

[42] In principle, there could be spillovers across a seller's multiple products. Appendix Table 5 examines such spillover effects on sales of peaches, the second most popular fruit consumed in the summer. The dependent variable is daily sales value in RMB (summed over all quality categories). We see no significant difference across the treatment groups.

With the evidence established above, we can now look at what happened to sellers' quality choices before and after the incentive was removed. Table 8 shows the difference-in-difference regression results. The even columns control for seller baseline characteristics. The coefficient for the interaction term between the incentive treatment and the post-incentive dummy is close to zero and non-significant for the laser group. On the contrary, sellers in the sticker incentive group reverted to a lower quality level after the incentive was removed. This is consistent with the discussion above that it may take a longer time to establish trust under stickers. Therefore, it is not clear how much the incentive has facilitated sellers initial reputation building within this short intervention.

5.5 A Qualitative Explanation for the Lack of Quality Differentiation at Baseline

Overall, the reduced form results are consistent with the model's predictions and provide a qualitative explanation for the lack of quality differentiation at baseline. Sales profits of the sticker group did not outperform the label-less group, which explains why no seller differentiated quality at sale at baseline despite the fact that stickers have long been cheaply available.[43] While sellers in the laser group earned higher sales profits, the relevant consideration is whether the increase in profits, netting out the effort costs of providing higher quality, justifies the fixed cost of the laser machine. One year after the intervention, when surveyors revisited these markets, none of the 57 sellers that could be tracked continued with quality differentiation. This suggests individual sellers would not have the incentive to take up this new technology themselves. The structural analysis will help to shed light on why that is the case, which is related to the structure of these markets.

Next, I estimate an empirical model of consumer learning and seller reputation to rationalize the experimental findings and use the structural model to perform counterfactual policy and welfare analysis.

6 An Empirical Model of Consumer Learning and Seller Reputation

The empirical model follows the same setup as the model outlined in Section 3. I first enrich the basic setup in Section 6.1. Estimation proceeds in two steps (Section 6.2). First, the dynamic demand model is estimated using the household panel data. Second, the supply-side parameters are estimated by solving for the sellers' optimal policies, taking the demand estimates as given. Section 6.3 discusses the results and examines model fit. Section 6.4 uses the structural estimates to simulate consumers' beliefs and sellers' net returns evolution under different branding technologies.

[43] In fact, the initial performance of sellers in the sticker group is very similar to those in the label-less group (see Appendix Table 4) though the latter reverted back to non-differentiation and the former continued to differentiate. This suggests that sellers may be in fact close to indifference between the two business strategies, which is consistent with the observed outcomes on sales profits.

6.1 Setup and Assumptions

6.1.1 Demand Side: A Model of Consumer Learning and Purchasing

The demand model follows the same basic setup as in Section 3. I start by restating the key assumptions and providing some qualitative justifications for these assumptions.

Assumption 1 (*Demand side*): (1) Consumers share common prior beliefs about the unobserved quality, which is believed to be fixed over time (for a given type of watermelon); (2) Consumers update only on the premium option; (3) Consumers make purchasing decisions to maximize current utility.

Assumption 1.1 is discussed in Section 3. Quality is operationalized as the probability of being good. Assumption 1.2 is consistent with the reduced form results in Panel B of Table 2 (see discussion in Section 5.1). Finally, the model abstracts away from forward-looking behavior (Assumption 1.3). As discussed in a recent review paper by Ching, Erdem, and Keane (2013), the literature on estimating consumer learning models has not reached a clear consensus on forward-looking versus myopic behavior. To model forward-looking behavior, one needs to solve a dynamic discrete choice problem. Besides the usual computational difficulties, the current setting imposes an additional challenge, which is that it may be hard to model the value of experimentation in the context of a new good as consumers' perceptions about future product availability, price and quality would matter. The goal of the empirical exercise is to estimate a parsimonious model that describes consumers' actual purchasing behavior, and that is also tractable enough to be integrated with the supply-side model.[44] As a first pass, given the seasonal nature of the fruit, if the option value of experimentation plays an important role, we would expect that the number of first-time buyers for the less-known premium product option to be higher in the initial period. However, there does not appear to be such a pattern in the data (Appendix Table 6).

Priors, distribution of outcomes, and updating

The prior distribution and the updating process are described in Section 3. Here, I enrich the setup by incorporating private experience shocks and an enlarged choice set that includes buying from other sellers. Let $e_{imjt} \in \{0, 1\}$ indicate whether a type j watermelon is satisfactory or not for individual i in market m at time t. There are three types 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 sources. Prior beliefs about the quality of the premium option are assumed to follow beta distribution with parameters (a_0, b_0). The posterior at time t is given by

[44] In practice, it is difficult to combine a complex dynamic demand system with a supply-side model (for example, see Ching (2010b) and Hendel and Nevo (2011)), and estimating a full dynamic game between forward-looking heterogeneous consumers and sellers under asymmetric information remains as an empirical challenge. Fershtman and Pakes (2012) propose an equilibrium concept, called the Experience Based Equilibrium, where players choose their optimal strategies based on their own observable experiences. The authors provide an estimation framework that is based on a reinforcement learning algorithm. In a similar spirit, one could view the beliefs evolution in this empirical model as consumers learning to converge to a steady state.

$(a_{im1t}, b_{im1t}) = (a_0 + s_{im1t}, b_0 + f_{im1t})$, where s_{im1t} and f_{im1t} are the numbers of satisfactory and non-satisfactory experiences individual i has had after time t. By Assumption 1.2, consumers do not update on the other options. Let q and $q + \Delta q$ denote the (degenerate) beliefs about the quality of other sources and the normal option. Δq captures any spillover effect (see discussion below).

Decision rule

Consumer's expected utility of purchasing option $j \in \{1, 2, 3\}$ at time t is

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

where $\mu_{imj,t-1}$ denotes consumer i's posterior for option j at the end of time $t - 1$. P_{mjt} is j's price in market m at time t. θ captures vertical taste differentiation, and is allowed to vary across consumers with different baseline self-reported willingness to pay for quality. The price coefficient α is allowed to be different for high- and low-income groups. Num_i is the number of watermelons consumed per week reported at baseline, which seeks to capture heterogeneous love for watermelons in general. η_i and ξ_i are unobserved preferences for the premium option and for the sample seller. For example, some consumers may have a predilection for expensive products, and some may be more likely to buy from a particular seller than from the others (i.e. horizontal taste differentiation). λ_m are market fixed effects, capturing time-invariant differences across markets. λ_t are time fixed effects, capturing aggregate time shocks that affect all markets, such as weather shocks.[45] ϵ_{imjt} are idiosyncratic random utility shocks, which are realized in each period before the purchasing decision is made. Let V_{imjt} denote the mean utility, excluding the random shock component.

There is an outside option with mean utility 0 for not purchasing any watermelon in a given period (denoted as $j = 0$).[46] Assuming risk neutrality, the consumer chooses j to solve

$$\max_{J=\{0,1,2,3\}} V_{imjt} + \epsilon_{imjt}$$

Further assuming that the idiosyncratic shocks ϵ_{imjt} follow i.i.d. Type 1 extreme value distribution, the choice probability takes a logit form:

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

[45] In estimation, I exclude the time fixed effect for the first period, thus the estimated time effects are relative to the first week. I estimate the full set of market fixed effects (as the utility specification does not contain a constant term).

[46] Like in all discrete choice models, the level and the scale are not identified. A common practice is to normalize the mean utility of the outside option and the variance of the error term.

6.1.2 The Supply Side: A Model of Seller Reputation Building

For the supply side, I focus on the laser groups, for which we have seen clear evidence for reputation building.[47] Sellers choose prices and quality to maximize the net present value of profits:

$$\max_{\{p_H^t, p_N^t, \gamma_H^t\}} \sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E}\Big(\left(p_H^t - p_W^t - C(\gamma_H^t)\right) Q_H^t + (p_N^t - p_W^t) Q_N^t + \mathbb{1}_{\text{Inc}} \times \phi(\gamma_H^t) B \Big) \tag{6}$$

$$\text{s.t.} \quad \{p_H^t, p_N^t\}, g(\mu^t, X) \to Q_H^t, Q_N^t \to_{\gamma_H^t} g(\mu^{t+1}, X), \quad g(\mu^0, X) \text{ given}$$

where $g(\mu^t, X)$ is the joint distribution of household beliefs (μ^t) and characteristics (X) included in the demand model, and it constitutes the seller's state variable. The prior mean $g(\mu^0, X)$ is given by the demand side, and the beliefs evolution is determined by the learning dynamics and seller's policies. In particular, the current prices (p_H^t and p_N^t) and the current state jointly determine the current sales (Q_H^t and Q_N^t), which, together with the current quality choice, determine the next period's state. A period is taken as a week to match the demand system.

The per-unit cost of the normal product is the wholesale price p_W, and the additional marginal cost of providing the premium product is parameterized as:

$$C(\gamma_H) = c \log\left(\frac{1-\underline{\gamma}}{1-\gamma_H}\right)$$

where $\underline{\gamma}$ denotes the average quality of the undifferentiated pool. $C(\gamma_H)$ captures the effort costs of sourcing better watermelons in the upstream (see discussion in Section 5.2). In the extreme case, if $\gamma_H = \underline{\gamma}$, the cost simply reduces to 0. Finally, the objective function for the incentive group contains an additional term $\phi(\gamma_H^t)B$ for each period, where B is the amount of the incentive (100 RMB) and $\phi(\gamma_H^t)$ is the probability of being rewarded. Since quality checks are conducted twice a week and the monetary reward is issued if the measured quality passes the pre-specified level on both tests, I assume $\phi(\gamma_H) = \gamma_H^2$ to match the empirical setting.[48]

The main estimation challenge for solving the dynamic optimization problem is that the state space is of infinite dimension. To make progress, I make the following important simplifying assumption:

Assumption 2 (*Supply side*): Seller pegs the normal pile price (p_N^t) at the market average in each period and chooses a once-for-all quality (γ_H) and markup (m_H) for the premium pile: $p_H^t = p_N^t + m_H$.

While this assumption is restrictive, it is consistent with sellers' actual behavior observed in the data. Appendix Figure 8 plots the price trajectories for the laser groups. On average, the normal pile price closely tracks the market average price; the latter slightly trends downward as the wholesale price

[47] Evidence for the sticker groups is mixed. See discussion in Section 5.

[48] The implicit assumption is that the pre-specified sweetness threshold matches consumer's subjective satisfaction assessment. Appendix Figure 7 plots the empirical CDF of the sweetness for the undifferentiated piles. 10.5 corresponds to the 73rd percentile of the distribution, which is close to the 30% empirical satisfaction rate in the household data for undifferentiated watermelons.

decreases over time. Importantly, the price difference between the premium pile and the normal pile remains quite stable over time. One concern is that the average results could mask underlying significant individual heterogeneity. Appendix Figure 9 looks at the pricing dynamics for a typical market. While we see some occasional price adjustments, those look sporadic and idiosyncratic rather than anything systematic. Appendix Figure 10 looks at quality dynamics (measured in sweetness) of the premium pile. Again, we do not observe any clear time pattern.[49] Appendix Table 10 examines the time dynamics for pricing and quality provision in a regression framework, and we see that the coefficients for the linear time variables are very close to zero.

The empirical patterns above provide qualitative justification for constraining the class of policies to once-for-all markup and quality. One explanation could be that frequent price adjustments may send some negative signals to consumers, and although quality differentiation happens daily, to actually fine-tune that to actual demand conditions may be hard and mentally costly. Having said that, one could imagine that as reputation is built up, a seller may well increase markup in longer-time horizons; the current model does not accommodate that possibility. Unfortunately, the data, which only lasts for 8 weeks, is limited in addressing these important long-term reputation dynamics. Given this limitation, the approach undertaken here searches for the optimal policies within the restricted class of policies.

6.2 Estimation Strategies and Identification

6.2.1 Demand Side: Simulated Maximum Likelihood Estimation

The demand model is estimated using simulated maximum likelihood.[50] I collapse the household panel purchasing data to household-week level and merge that with the market-week level average prices calculated from the surveyors' records.[51] For each purchase experience, the household reports a satisfaction rating from 1 to 5. I recode 5 to be satisfactory and $\{1, 2, 3, 4\}$ as well as missing values to be non-satisfactory.[52] To allow prior beliefs to be different under different branding technologies, I allow a_0 and b_0 to depend on laser and sticker. In other words, we can think of households living

[49] There appears to be discrete jump in quality for the incentive group after the first week. This could be due to sellers having initial doubts about receiving the monetary rewards at the beginning of the intervention.

[50] Train (2009) provides a detailed exposition on estimating mixed logit model using panel data.

[51] There are a few occasions with multiple purchases within a week. I recode multiple purchases from different sources as purchasing from the mode source and recode the number with a purchase dummy to fit the discrete choice framework. Issues with treating non-recording as non-purchase are discussed in Appendix C.2.

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

in different markets as facing different choice sets: households in the laser and sticker markets face a premium option with either a laser or a sticker label. To simplify the analysis, for households in the label-less markets, I restrict the data to week 3 and onward after most label-less sellers reverted back to non-differentiation. Therefore, we can think of these households as facing a restricted choice set without the premium option. Finally, to allow for spillover effects across a seller's multiple products, I estimate separate belief shifters, $\Delta q(s)$ and $\Delta q(l)$, for the normal option for sellers in the sticker and laser groups.

The likelihood of household i for making an observed sequence of purchases can be computed as:

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

Assuming that the random effects η and ξ follow independent distributions of $\log(\mathcal{N}(m(\eta),\sigma(\eta)))$ and $\mathcal{N}(m(\xi),\sigma(\xi))$, the average likelihood function for each household, $\tilde{l}_i$, can be computed by averaging l_{ni} over a large number of draws. The final objective function is obtained by multiplying $\tilde{l}_i$ across all households and taking log. Standard errors are computed using the outer product of gradient estimate of the asymptotic covariance matrix. Details for the estimation procedure and standard error calculation are provided in Appendix D.1.

The identifying assumptions are twofold. First, market and time fixed effects fully capture unobserved time-varying shocks that directly affect both prices and demands for a market.[53] Second, η and ξ fully capture unobserved persistent individual heterogeneity. Under these two assumption, with one period data on market shares, we can identify the market specific constants, the mean of the prior beliefs multiplied by the vertical taste parameters, the price coefficients, the coefficient for *Num*, and the distributions of η and ξ (following standard arguments in the discrete choice literature). Parameters θ, a_0 and b_0 are identified from the dynamic purchasing patterns. Intuitively speaking, if repurchasing decisions are very responsive to past experiences, it could because households either care a lot about quality (large θ) or the variance of the prior is large (small a_0 and b_0). However, the *difference* in the *change* in the repurchasing probabilities between going from zero to one good (or bad) experience and that going from one to two separately identifies these parameters. In particular, the difference should be bigger under the large variance story than it is under the large willingness to pay story because belief updating is more salient for the former case.[54]

[53]The identifying assumptions will be violated if there are unobserved time-varying market-specific factors that affect both demands and prices. One primary concern is information diffusion: sellers may raise prices following a period of good behavior, and the resulting optimistic beliefs could be disseminated among the local population through word-of-mouth and thus directly affect an individual's demand other than through the individual's own beliefs. I address this concern more in Section 6.3.

[54]Appendix Table 7 summarizes the repurchasing probabilities conditioning 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. For households in the laser group, going from zero experience to one satisfactory experience increases the purchasing probability by about 63%, but having one additional satisfactory experience further increases the probability

6.2.2 Supply Side: Minimum Distance Estimator

Taking the demand estimates as given, the supply side parameters are estimated using a minimum distance estimator. Ideally, one would like to solve for the optimal policies market by market and apply the minimum distance estimator to the full vector of policies for all sellers. Unfortunately, γ_H is not observed for each individual seller and cannot be reliably approximated using the empirical satisfaction rate in the household data due to the small sample size for each market.[55] Given this data limitation, I first construct a hypothetical *average market* by pooling together all households in the laser markets and averaging the fixed effect estimates for these markets. There are 194 households in the sample; however, in reality, sellers face a much larger market size, thus I scale up the market size by 4.5 to match the initial period's total sales quantity. I then solve for the optimal policies, m_H^* and γ_H^*, for a seller facing this hypothetical *average market*. The structural parameters are estimated by minimizing the distance between the optimal policies and the empirical average policies:

$$v(\delta, c) = \sum_{g \in \{\textbf{laser inc,laser non-inc}\}} \left(\gamma^*_{Hg}(\delta, c) - \overline{\gamma}_{Hg}\right)^2 + \left(m^*_{Hg}(\delta, c) - \overline{m}_{Hg}\right)^2 \tag{7}$$

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

One concern of looking just at the average behavior is that the average could mask significant individual heterogeneity across sellers. Appendix Table 12 and 13 examine sellers' markup and quality choices. Generally speaking, while policies do vary across sellers, most seem to be on dimensions related to the demand conditions, which are already captured in the current framework. Having said that, there could be other important dimensions of individual heterogeneity that are not observed in the data but that affect a seller's reputational incentive. The current approach focuses on the effects of the demand conditions and therefore abstracts away from other aspects of individual heterogeneity.

For each given set of δ and c, the optimal policies are found using grid search. The objective function

by only 3.5%. This pattern indicates a very noisy prior. However, the fraction of repurchasing also goes up with one bad experience. This is not surprising given that the compositions of households are different for the different cells. Nonetheless, the difference in the repurchasing probabilities under (0,1) and (1,0) can be interpreted as the effect of learning because the total number of experiences is the same in these two cases, which controls for the composition effect. The fact that this difference is much more pronounced under laser than under sticker is consistent with the reduced form results in Table 2.

[55] An alternative approach would be to estimate γ_H together with the other structural parameters. Intuitively speaking, a seller's sales trajectory is informative about her underlying quality provision. However, leaving aside the computational burden, the problem with this approach is that the demand system is only estimated for half of the markets where the household data are collected. Even for those, the market fixed effects are estimated on a small sample of roughly 20 households per market, hence the estimates can be quite imprecise. This issue is reflected in the large standard errors for the fixed effect estimates shown in Appendix Table 8.

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

6.3 Results and Model Fit

The simulated ML estimates are presented in Table 9. Market and time fixed effect estimates are abbreviated from the main table and are reported in Appendix Table 8. If I estimate the model unconstrained, a_0 turns out to be slightly negative. This is because of the small market shares of the premium option observed in the data.[56] In other words, the data suggests a very pessimistic prior as viewed through the lens of this model. Given the beta prior distribution, I constrain a_0 to be non-negative in the estimation. Estimates of the key parameters are qualitatively similar to the unconstrained case.

Looking at column 1, the estimate for b_0 is 0.938 for laser and 2.578 for sticker. The point estimates are consistent with the discussion in Section 4.2 and suggest that the prior beliefs are more *stubborn* under sticker than under laser. In particular, one satisfactory experience updates the posterior mean to 0.52 under laser, but only to 0.28 under sticker.

Beliefs about the quality of the undifferentiated option from the other sellers is estimated to be 0.307. This number matches well with the 30% empirical satisfaction rate in the household data. The negative $\Delta q(s)$ suggests that consumers in the sticker markets seem to perceive the normal pile as having lower quality if sellers sell it beside another pile that is labeled with a sticker and that is purported to be of a higher quality.[57] The signs of the other estimates are consistent with expectations.

To interpret the magnitudes and check that the point estimates are plausible, I calculate the price elasticity faced by sellers in the label-less group (for selling the undifferentiated pile).[58] The price elasticity averaged over all the label-less markets for this time period turns out to be -2.14.[59] Using the

[56]In principle, including a product-specific constant (in this case, the random effects η and ξ) in the utility specification could help to alleviate the constraint. However, it doesn't in this case–the estimate of a_0 hits the zero boundary regardless of whether we include the random effects or not. This is because the data also points towards a very small a_0 to match the fast switching response: consumers with 1 or 0 good experience display very different repurchasing behavior; the largest possible difference the model could allow is when $a_0 = 0$, together with a small b_0. It's clear from this discussion that identification of the prior parameters relies on the dynamic purchasing patterns, which can be demanding given the short panel we have. An alternative approach is to constrain the prior mean to be the same as the existing option (q) and estimate the sum of $a_0 + b_0$, the smaller the value the larger the variance. The results (not shown) again indicate a much more stubborn prior under sticker than under laser. However, a simple likelihood ratio test strongly rejects the alternative model at 1% level (against the baseline model in Column 1 of Table 9).

[57]While the underlying psychology is not clear, such beliefs do seem to be justified in light of sellers' actual behavior: without spending extra effort sourcing better watermelons in the upstream, the quality of the two piles would indeed be negatively correlated. As shown in column 4 of Table 4, on average the quality of the normal pile for the sticker group appears lower than the market average.

[58]For sellers who differentiate quality at sale, changing price in one period affects not only the current sales but also future sales by affecting the next period's beliefs. I investigate the dynamic price elasticities in Appendix D.2.

[59]For the random effects, 300 Halton draws are taken and the results are averaged. See Appendix D.2 for details.

Lerner's rule, this implies a markup of 47% for profit maximizing. This number is quite close to the average empirical markup, which is around 43%. To further examine goodness of fit, I compare the key moments in the actual data versus those in the simulated data using the model estimates. Details on the simulation procedure are presented in Appendix D.3. Appendix Figure 5 shows that the predicted market share for each product option matches quite well with the empirical share. Appendix Figure 6 looks at the repurchasing probabilities conditioning on various experience combinations. Overall, the fit looks reasonable: the purchasing patterns generated by the prior estimates and the Bayesian learning process mimic the actual purchasing patterns well.

Columns 2 to 4 of Table 9 consider three extensions to the baseline model as robustness checks. Column 2 includes a product-specific constant ν for the premium option under laser label to account for any direct utility of laser. Column 3 incorporates reputation spillover via correlated learning by allowing the posterior for the premium pile to enter linearly into the mean utility of the normal pile (i.e. good experiences from the premium pile may lead consumers to favor the sample seller in general).[60] Column 4 includes a linear function of the market average beliefs (computed as the average beliefs of households in a given market at a given time) in the mean utility of the premium option to control for information diffusion. While the measures and the approaches are not perfect, the results are reassuring. Overall the ML estimates stay quite robust across various specifications and the likelihood ratio test does not reject the baseline model. Therefore, in what follows, I shall stick with the baseline estimates in column 1 for examining the seller's problem and calculating welfare.

Taking the demand estimates in column 1 of Table 9, δ and c are estimated to be 0.98 and 0.64. Appendix Figure 11 plots the value of the objective function as δ and c vary and Appendix Table 11 reports the optimal policies under various δ and c in comparison with the empirical policies. We see that the model is able to generate a quality gap between the incentive and non-incentive groups (0.48 versus 0.41), which is fairly close to the empirical gap (0.53 versus 0.40). The optimal markup for the incentive group is also higher than that for the non-incentive group as they are in the data, though the magnitudes are larger than the empirical values. To check that the demand estimates are plausible, I simulate aggregate sales outcomes using the parameter estimates and the average empirical policies (see Appendix D.3 for details of the simulation procedure). Results are shown in Table 10. Overall, the simulated weekly average sales quantity and profits are in line with the actual sales outcomes as calculated from seller's data (averaged across all sellers in a given group). In particular, we have seen in the reduced form result that the actual sales of the premium pile for the incentive group outperform those for the non-incentive group (Table 6). Here, we see that the dynamic demand system, combined with the higher quality provision for the incentive group, is able to generate that result.

[60] To formally model correlated learning, one would estimate a model that is analogous to nested logit in static discrete choice where consumers first choose seller and then pick pile. However, in this setting, the local markets are quite compact, so stores are located close to one another. Thus it is reasonable to think that consumers make purchase decisions at the product level rather than in two steps. This could explain why the estimated spillover effect through correlated learning (column 3) appears to be small, compared to the direct effects, s_0 and s_1.

6.4 Beliefs and Net Returns Evolution

I now use the structural estimates to examine how underlying beliefs endogenously evolve over time, and how prior beliefs could affect seller's reputational incentive.

Panel A of Figure 6 plots the group average beliefs evolution for the quality of the premium pile. In particular, I take the demand estimates in column 1 of Table 9 and feed them through the actual purchasing and experience realizations to compute the posterior for each household in each period. I then average that across all households in a given treatment group to get the group average beliefs, which can be seen as a proxy for reputation. First, we see that the average beliefs are the highest for the laser incentive group by the end of the intervention, consistent with the reduced form result on households' endline perceptions (Table 7). Next, conditioning on the incentive treatment, average beliefs rise faster under laser than under sticker, in line with the reduced form patterns in Table 2.

The average beliefs evolution is a result of two underlying effects: first, laser branding induces faster belief updating; second, laser branding induces sellers to provide higher quality, resulting in more satisfactory experiences. To decompose the two effects, I take the sample of households in the sticker group and first simulate the average beliefs evolution under the sticker group's empirical policy. The result is shown by the dashed line in Panel B. Next, I plot the counterfactual beliefs evolution under the same policy but replace the prior beliefs parameters with those for laser. The result is shown by the dotted line with the square markers. Finally, I plot the counterfactual beliefs evolution when both the empirical policy and the prior beliefs are replaced with those of the laser group. The result is shown by the dotted line with the diamond markers. We see that holding the supply-side behavior as fixed, laser branding alone has a significant impact on beliefs evolution as shown by the enlarging gap between the first two lines. The different prior beliefs shape sellers' reputational incentives, which further drive markets to different outcomes over time. The gap between the dotted line with the diamond markers and the dashed line represents the total effect.

Next, I simulate sellers' net profits evolution under sticker and laser, under the empirical policy of the laser non-incentive group. The results are shown in columns 1 and 3 of Table 11. In column 4, I simulate the baseline case with no differentiation as a comparison. We see that the two-season discounted sum of net profits, taking into account the effort costs, is 51% higher under laser than under non-differentiation; on the other hand, there is not much difference between quality differentiation under sticker and the baseline. Figure 7 plots the net profits evolution. An extrapolation to 5 seasons suggests that there might be large gains under laser: the five-season discounted sum of net profits is $\sim$ 13 kRMB higher than baseline ($\sim$ 11 kRMB). However, this increase is still not large enough to justify the initial investment cost of the laser machine, which is on the order of 50-60k RMB. There could be two reasons for this: first, each individual seller's market size is small; and second, it may be difficult for sellers to extract all the consumer surplus. The former indicates a collective action failure since one laser machine can serve multiple markets and the total gain in producer surplus can exceed the costs. The latter points to the importance of understanding the role of fragmented markets in the presence of information problems. The next section conducts counterfactual exercises to shed light on these interactions and examine the welfare effects of government policies in regulating markets with both information frictions and imperfect competition.

7 Welfare Effects of Information Frictions and Fragmented Markets

In a second-best world with multiple frictions, the welfare implication of each friction is theoretically ambiguous as the different frictions could counteract. In particular, while market power generally distorts quality provision from the first best (which is known as the Spence distortion), it also internalizes the return of investing in reputation by allowing sellers to capture a larger portion of the gain in consumer surplus. The latter channel could encourage quality provision in markets with information problems, where quality cannot be directly enforced. To examine the interaction and highlight the tradeoffs faced by policy makers, I conduct counterfactual exercises that remove one imperfection at a time in order to isolate the effect of the other. The results are presented in Table 11.

For all the exercises, I calculate the welfare for the same hypothetical *average market* as described in Section 6.2. Therefore, the numbers are comparable to those shown in Table 10. The key outcomes of interest are consumer surplus, producer surplus (sales profits net of effort costs) for both the sample seller and the other sellers, and total surplus. The numbers reported are five-season discounted sums.

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

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

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

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

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

With information problems, consumer surplus takes a more complicated form because beliefs under which purchasing decisions are made are different from the truth. Leggett (2002) develops a solution to this problem for Type-I extreme value random utility errors with constant marginal utility of wealth.

In particular, for consumer i in a given period t, the expected maximum utility (in RMB) is given by:

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

where

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

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

The baseline benchmark. Column 1 of Table 11 calculates the welfare for the baseline scenario with no quality differentiation. Using column 1 as the benchmark, I next calculate the counterfactual outcomes without information frictions. That is, for any quality that a seller chooses, she could perfectly convey that information to consumers at the point of transaction.

Symmetric information: one seller deviation. Column 2 considers a single seller deviation. I first solve for the seller's optimal quality and markup for the premium pile, holding price of the normal pile and the other sellers' strategy the same as in column 1. The optimal quality of the premium pile turns out to be 0.769, much higher than that of the normal pile, and the optimal markup is 1.156 RMB/Jin, which is also quite high given that the average price of the normal pile is 0.98 RMB/Jin over this period. Under these policies, the seller's net profit is almost 7 times higher than the baseline case. This result demonstrates that without information frictions, baseline cannot be an equilibrium because there is a large single-seller profitable deviation.

Symmetric information: separating equilibrium. Column 3 computes the equilibrium outcome under symmetric information.[61] For each γ_H and m_H chosen by the other sellers, I first solve for the optimal γ_H^* and m_H^* of the sample seller. A symmetric Nash equilibrium is found by searching for the fixed point. Here and in subsequent analyses, I focus on the best equilibrium for sellers (i.e. maximizing net profits) in case of multiple equilibria. We see that competition puts a downward pressure on price and increases quality. Consumer surplus is significantly higher than that in column 2 because of the lowered price and enlarged choice set. A comparison of the total surplus in columns 1 and 3 shows that information frictions result in a welfare loss of about 66.4%.

[61] For this and all subsequent analysis with market competition, I consider a hypothetical duopoly world of the *average* sample seller and a *representative average competitor* by averaging the market fixed effect estimates in the demand model. The latter, however, is a very crude proxy because the number of competitors varies by market, and therefore the fixed effect estimate, which can be interpreted as the inclusive value of a nest consisting of all other sellers, needs to be adjusted before taking the average. If we are willing to assume homogeneity across the other sellers, then knowing the actual number of competitors in each market, we could recover the fixed effect of a single competitor from the market fixed effect.

Symmetric information: first best. Column 4 solves for the first-best outcome. As studied in Spence (1975), the quality choice of a monopolist (or oligopolists) could be either higher or lower than the socially efficient level because the former considers the effect on the marginal buyer whereas the social planner cares about the average buyer. Because of that, the welfare implication of information frictions is theoretically ambiguous. However, a comparison of columns 3 and 4 shows that quality under imperfect competition is lower than the first-best level. Therefore, by further distorting quality downwards, the information problem would unambiguously exacerbate the welfare loss in this setting.

Comparing columns 1, 3 and 4, we can see that the welfare loss caused by market power (column 3 versus column 4) is small relative to that caused by the information problem (column 1 versus column 3) in this setting. One reason is that these markets are already quite competitive at baseline. Next, I calculate the welfare under asymmetric information and examine the role of market competition.

Asymmetric information: one seller. The middle panels of columns 5 and 6 calculate the welfare under the hypothetical scenarios where one seller is induced to provide quality with the new branding technology, with and without the incentive, and consumers' beliefs have converged to the actual quality choices. Compared to column 1, total surplus is 42.6% higher without the incentive and 70.7% higher with the incentive. The bulk of the welfare gain comes from increases in consumer surplus. Figure 8 plots the distributions of consumer surplus under the baseline case, the two hypothetical scenarios and the first best. Consumers are strictly better off with the premium option because their choice set is enlarged, and high-valuation consumers could self-sort into buying higher quality, albeit more expensive, product.

Having said that, the above numbers are only suggestive because it may take a very long time for beliefs to actually converge. The bottom panels of columns 5 and 6 compute the discounted sum of surpluses, taking into account the adjustment process. Compared to column 1, the increases in total surplus are 49k and 65k RMB for the non-incentive and incentive cases respectively. These gains are on the same order of magnitude with the cost of laser machines. While an individual seller would not undertake such an investment, a third-party could invest in the new technology and subsidize it for the sellers to enhance societal welfare.

Further comparing the outcomes in columns 5 and 6, we see that the difference in the discounted sum of total surplus (16,256 RMB) exceeds the amount of the initial incentive, which is only 600 RMB (100 RMB per week for six weeks). As shown in Section 5.4, the quality gap between the laser incentive and non-incentive groups appears to persist beyond the incentive period. If we take that seriously and are willing to extrapolate, the result suggests that providing sellers with a temporary monetary incentive to facilitate the initial reputation building could raise welfare.

Asymmetric information: market competition. To examine the effect of market competition in the presence of information asymmetry, I compute the symmetric Nash equilibrium outcome when all sellers in a market are given access to the new technology and simultaneously choose once-for-all quality and markup. The result is shown in column 7.[62] We see that competition induces sellers to provide higher quality (compare to the monopoly case in column 5); however, quality is still quite low compared to the first-best level (column 7 versus 4). One reason is that market competition on price could discourage quality provision.

Suppose government could regulate the price for the premium product and would still let sellers compete on the quality dimension. How much could this increase welfare? We can think of this as analogous to the first-best case but under asymmetric information when it is impossible to directly enforce quality. The result is shown in the last column of Table 11. For each markup level set by the social planner, I first search for the best symmetric equilibrium for sellers; the optimal markup is then found by maximizing the discounted sum of total surplus. In line with the discussion above, the social planner would want to set a higher markup to ease competition, which leads to higher quality provision than the case where sellers compete on both price and quality (column 7). That being said, after taking the adjustment process into account, the additional welfare gain of regulating price (column 8 versus 7) is small. The reason is that higher markup directly discourages sales, and thus beliefs take an even longer time to take off. On the other hand, lowering the markup would result in lower effort. This is a key tradeoff that the government faces in regulating markets with information problems.

One important caveat with this counterfactual exercise is that consumers' learning dynamics for the premium option are held the same under the case when it is provided by all sellers in the market and by a single seller. In reality, the learning dynamics may be quite different under these two scenarios. For example, learning may be correlated across similar-looking products, and if so, there would be reputation spillovers across sellers. One seller's reputation is adversely affected if the others do not behave well; anticipating that, sellers may further reduce their quality. This channel is distinct from the channel of price competition, which is what the counterfactual exercise addresses. Next, I shall conclude with some potential avenues for exploring this issue in future research.

8 Conclusion

This study empirically examines the dynamic interactions between sellers and consumers in an experience good market setting, the local watermelon markets in China. I find that information frictions and fragmented markets lead to significant under-provision of quality in this setting. Though there is a high demand for quality, trust could take a long time to establish under the existing branding technology, which makes reputation building a low return investment. While there is a new branding technology that could enhance consumer learning, small individual sellers do not have the incentive to invest in this technology due to their small market size and intense competition. An empirical model of

[62] There is another low-quality equilibrium with $\gamma_H^* = 0.4$ and $m^* = 0.12$. Outcomes under this equilibrium are presented in Appendix Table 14, in comparison with the high equilibrium.

consumer learning and seller reputation reconciles the experimental findings and quantitatively explains the absence of quality differentiation at baseline. Finally, third-party interventions that subsidize initial reputation building for sellers could enhance welfare. The result can be interpreted as suggesting that there may be a profitable entry opportunity for a large upstream firm that can employ the new branding technology and build a reputation for quality over time.

Though the exact learning processes and reputation dynamics are different for different goods,[63] the broad takeaway from this study is that prior beliefs and consumer learning matter. Good reputations may take a long time to establish, as is the case with the Wholefoods brand in the United States. In developing countries that lack such reputable entities, people's current beliefs are what matters. Rampant counterfeiting activities under the old sticker technology created distrust among consumers. Therefore, developing new branding technologies that are effective at deterring counterfeits and establishing trust can be crucial in firms' incentive of inventing and introducing new high-quality products.

Given the short intervention period, I stress that the welfare and counterfactual analysis hinges on several important assumptions about long run reputation dynamics and market environment. Nonetheless, the framework provides a first step upon which a longer-term study can be based to shed light on many interesting remaining questions. For example, it is possible that as a seller establishes a reputation for high quality under the expensive branding technology, counterfeiters may enter the market *if* the profits earned during the process of destroying the brand exceeds the initial fixed cost. This may be precisely why sellers choose not to invest in quality under the old technology, which is cheap to fabricate. While the expensiveness of laser branding potentially deters future entry of counterfeiters, whether that would be the case or not is an open question, which also depends on consumer learning dynamics in such circumstances.

The counterfactual analysis also assumes a fixed upstream supply. However, it is possible that as reputation is built up in the downstream, the incentive may trickle up to motive higher efforts to grow watermelons in the upstream. In the short run, with fixed upstream supply, the welfare gain from downstream differentiation would purely arise from allocative efficiency; however, the gain may be larger in the long run with endogenous upstream response. In fact, one of the largest watermelon seed companies in China, Hebei Shuangxing Seed Co., Ltd., is starting a new business venture to contract with farmers, invest in high quality production and build a premium brand using the laser technology.

Finally, the current study is limited in its investigation of market dynamics wherein multiple sellers sell the same newly introduced experience goods. In this intervention, only one seller in each local market was treated; the others did not strategically respond because they were not given access to the new technology. One could imagine that in other realistic settings, markets could feature both asymmetric

[63] For example, for some other experience goods such as milk powder, learning is mostly via catastrophic news and scandals. The reputation dynamics in those markets are likely to be very different from that for watermelons. Recent theoretical work by Board and Meyer-ter Vehn (2013) distinguishes the case of good- versus bad-news learning and provides some guidance for future empirical work.

information and oligopolistic competition (see Villas-Boas (2004) for a theoretical discussion). One of the counterfactual exercises considers a case where all sellers are given access to the new technology and compete for demand. The result highlights the importance to understand the effects market competition in the presence of information problems. To better understand sellers' reputational incentives and their quality provision decisions in settings with both asymmetric information and oligopolistic competition, one needs better knowledge of consumer learning dynamics in such circumstances. A possible extension to the current study is to treat multiple sellers in a market. I leave that as a potential avenue for exploration in the future.

References

Ackerberg, D. A. (2003): "Advertising, learning, and consumer choice in experience good markets: an empirical examination*," International Economic Review, 44(3), 1007–1040.

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

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

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

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

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

Board, S., and M. Meyer-ter Vehn (2013): "Reputation for quality," Econometrica: Journal of the Econometric Society, 81(6), 2381–2462.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014): "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh," Econometrica: Journal of the Econometric Society, 82(5), 1671–1748.

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

Ching, A., C. R. Clark, I. J. Horstmann, and H. Lim (2011): "The Effects of Publicity on Demand: The Case of Anti-Cholesterol Drugs," SSRN Working Paper Series.

Ching, A., and M. Ishihara (2010): "The effects of detailing on prescribing decisions under quality uncertainty," Quantitative Marketing and Economics, 8(2), 123–165.

Ching, A. T. (2010a): "Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration," International Journal of Industrial Organization, 28(6), 619–638.

——— (2010b): "A dynamic oligopoly structural model for the prescription drug market after patent expiration," International Economic Review, 51(4), 1175–1207.

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

Chintagunta, P. K., R. Jiang, and G. Z. Jin (2009): "Information, learning, and drug diffusion: The case of Cox-2 inhibitors," QME, 7(4), 399–443.

Coscelli, A., and M. Shum (2004): "An empirical model of learning and patient spillovers in new drug entry," Journal of Econometrics, 122(2), 213–246.

Crawford, G. S., and M. Shum (2005): "Uncertainty and learning in pharmaceutical demand," Econometrica, 73(4), 1137–1173.

Dellarocas, C. (2006): "Reputation mechanisms," Handbook on Economics and Information Systems, pp. 629–660.

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

Duflo, E., M. Kremer, and J. Robinson (2011): "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," American Economic Review, 101, 2350–2390.

Dupas, P. (2014): "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment," Econometrica: Journal of the Econometric Society, 82(1), 197.

Erdem, T., and M. P. Keane (1996): "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets," Marketing science, 15(1), 1–20.

Erdem, T., M. P. Keane, and B. Sun (2008): "A dynamic model of brand choice when price and advertising signal product quality," Marketing Science, 27(6), 1111–1125.

Fershtman, C., and A. Pakes (2012): "Dynamic games with asymmetric information: A framework for empirical work," The Quarterly Journal of Economics, 127(4), 1611–1661.

Fischer, G. (2013): "Contract Structure, Risk-Sharing, and Investment Choice," Econometrica, 81(3), 883–939.

Foster, A. D., and M. R. Rosenzweig (2010): "Microeconomics of technology adoption," Annual Review of Economics, 2(1), 395–424.

Fudenberg, D., D. Kreps, et al. (1988): "Learning, experimentation, and equilibrium in games," photocopy, Department of Economics, Stanford University.

Fudenberg, D., and D. K. Levine (1993): "Self-confirming equilibrium," Econometrica: Journal of the Econometric Society, 61(3), 523–545.

Hausmann, R., and D. Rodrik (2003): "Economic development as self-discovery," Journal of development Economics, 72(2), 603–633.

Hendel, I., and A. Nevo (2011): "Intertemporal price discrimination in storable goods markets," Discussion paper, National Bureau of Economic Research.

Hertzendorf, M. N. (1993): "I'm not a High-Quality Firm-but I Play One on TV," The Rand Journal of Economics, pp. 236–247.

Homer, P. M. (1995): "Ad size as an indicator of perceived advertising costs and effort: The effects on memory and perceptions," Journal of Advertising, 24(4), 1–12.

Horstmann, I. J., and G. M. MacDonald (1994): "When is advertising a signal of product quality?," Journal of Economics & Management Strategy, 3(3), 561–584.

Hubbard, T. N. (2002): "How Do Consumers Motivate Experts? Reputational Incentives in an Auto Repair Market*," Journal of Law and Economics, 45(2), 437–468.

Huffman, W. E., M. Rousu, J. F. Shogren, and A. Tegene (2007): "The effects of prior beliefs and learning on consumers' acceptance of genetically modified foods," Journal of Economic Behavior & Organization, 63(1), 193–206.

Israel, M. (2005): "Services as experience goods: An empirical examination of consumer learning in automobile insurance," American Economic Review, 95(5), 1444–1463.

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

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

Kalra, A., S. Li, and W. Zhang (2011): "Understanding responses to contradictory information about products," Marketing Science, 30(6), 1098–1114.

Kandori, M. (2008): "Repeated games," New Palgrave Dictionary of Economics, 2nd edition, Palgrave Macmillan.

Keller, K. L., and D. R. Lehmann (2006): "Brands and branding: Research findings and future priorities," Marketing science, 25(6), 740–759.

Kihlstrom, R. E., and M. H. Riordan (1984): "Advertising as a Signal," The Journal of Political Economy, pp. 427–450.

Kirmani, A. (1990): "The effect of perceived advertising costs on brand perceptions," Journal of Consumer Research, pp. 160–171.

——— (1997): "Advertising repetition as a signal of quality: If it's advertised so much, something must be wrong," Journal of advertising, pp. 77–86.

Kirmani, A., and P. Wright (1989): "Money talks: Perceived advertising expense and expected product quality," Journal of Consumer Research, pp. 344–353.

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

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

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

Lewis, R. A., and D. H. Reiley (2011): "Does Retail Advertising Work?," .

Linnemer, L. (2002): "Price and advertising as signals of quality when some consumers are informed," International Journal of Industrial Organization, 20(7), 931–947.

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," .

Maheswaran, D., D. M. Mackie, and S. Chaiken (1992): "Brand name as a heuristic cue: The effects of task importance and expectancy confirmation on consumer judgments," Journal of consumer psychology, 1(4), 317–336.

Mailath, G. J., and L. Samuelson (2013): Reputations in repeated games, vol. 4.

McMillan, J., and C. Woodruff (1999): "Interfirm relationships and informal credit in Vietnam," Quarterly journal of Economics, pp. 1285–1320.

MILGROM, P., AND J. ROBERTS (1986): "Price and advertising signals of product quality," The Journal of Political Economy, 94(4), 796–821.

MUNSHI, K. (2004): "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," Journal of development Economics, 73(1), 185–213.

NELSON, P. (1970): "Information and consumer behavior," The Journal of Political Economy, 78(2), 311–329.

OSBORNE, M. (2005): "Consumer Learning, Habit Formation and Heterogeneity: A Structural Examination," Manuscript, Univ. Chicago.

PARK, C. W., B. J. JAWORSKI, AND D. J. MACLNNIS (1986): "Strategic brand concept-image management," The Journal of Marketing, pp. 135–145.

PRASAD, K., AND C. S. DEV (2000): "Managing hotel brand equity: A customer-centric framework for assessing performance," The Cornell Hotel and Restaurant Administration Quarterly, 41(3), 22–4.

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

ROBERTS, J. H., AND G. L. URBAN (1988): "Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice," Management Science, 34(2), 167–185.

SCHMALENSEE, R. (1978): "A model of advertising and product quality," The Journal of Political Economy, pp. 485–503.

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

SIMESTER, D. (2015): "Field Experiments in Marketing," Manuscript, MIT Sloan School of Management.

SPENCE, A. M. (1975): "Monopoly, quality, and regulation," The Bell Journal of Economics, 6(2), 417–429.

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

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

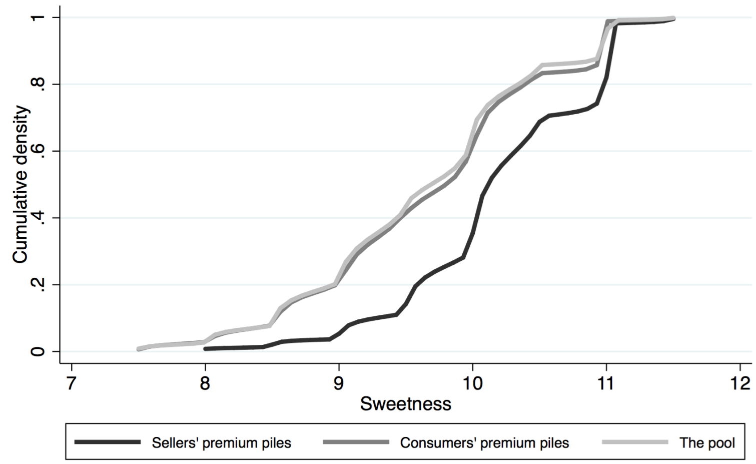
VILLAS-BOAS, J. M. (2004): "Consumer learning, brand loyalty, and competition," Marketing Science, 23(1), 134–145.

YI-LIN TSAI, C. S. D., AND P. CHINTAGUNTA (2015): "What's In A Name? Assessing the impact of rebranding in the hospitality industry," Journal of Marketing Research.

ZHANG, Y., ET AL. (2015): "The Impact of Brand Image on Consumer Behavior: A Literature Review," Open Journal of Business and Management, 3(01), 58.

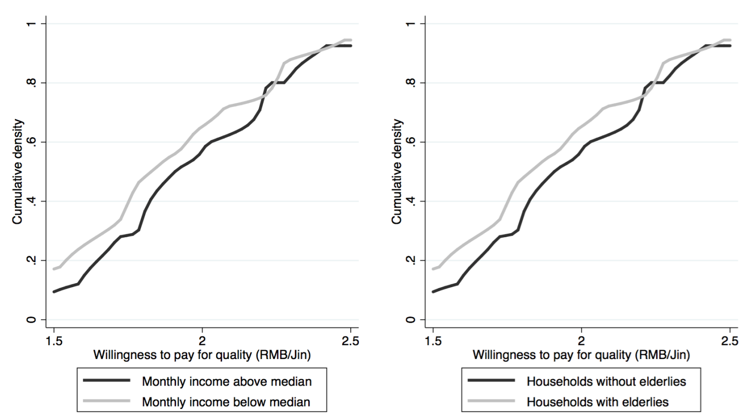
ZHAO, H. (2000): "Raising awareness and signaling quality to uninformed consumers: A price-advertising model," Marketing Science, 19(4), 390–396.

Figure 1: Asymmetric Information Between Sellers and Consumers in the Watermelon Market



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

Figure 2: Heterogeneity in Consumers' Willingness to Pay for Quality



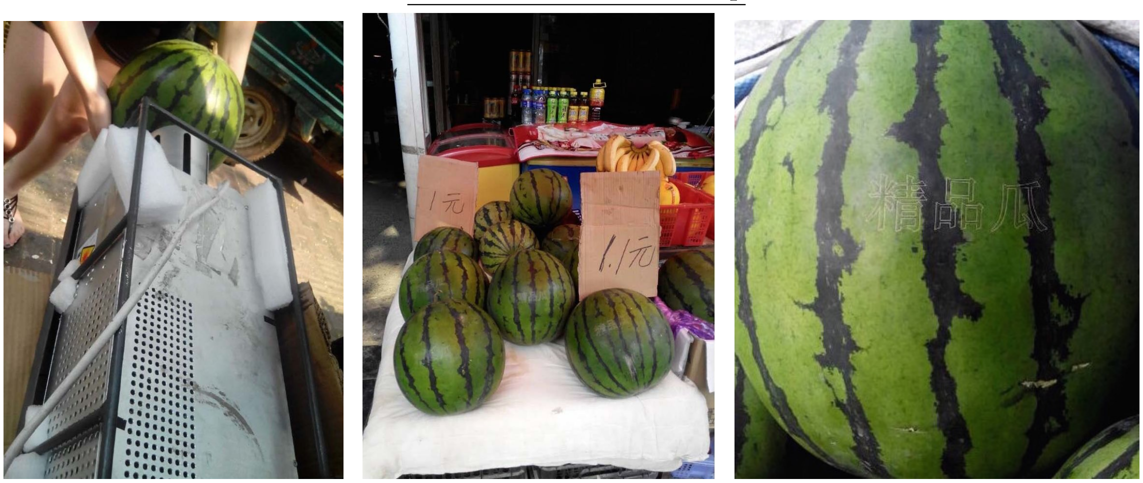
Note: This figure shows the heterogeneity of households' self-reported willingness to pay for quality elicited in the baseline survey. Households were asked to consider a hypothetical situation where they see two piles of watermelons sold in the local market, one pile of ordinary quality at 1.5 RMB/Jin 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) were 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 cumulative distributions for households with monthly income above and below the median. The right figure shows the distributions for households with and without elderly members.

Figure 3: Pictures of the Branding Treatments

Panel A. The Label-less Group

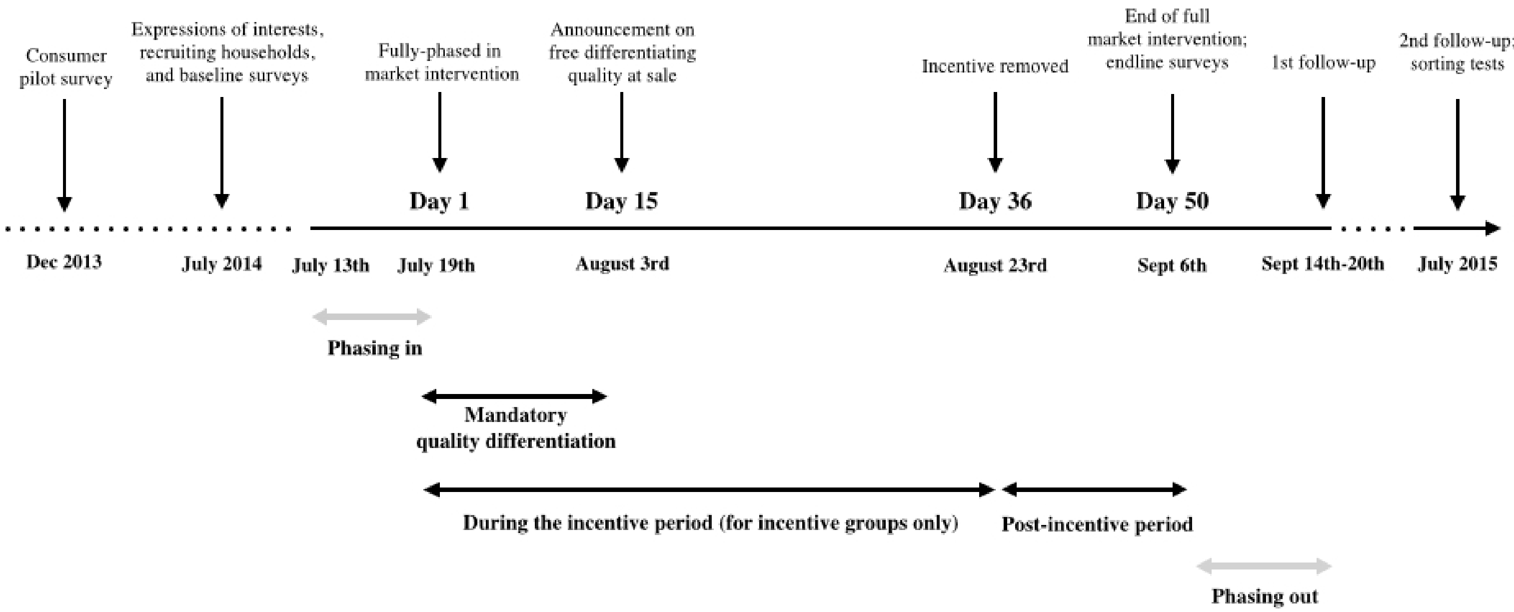
Panel B. The Sticker Group

Panel C. The Laser Group



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

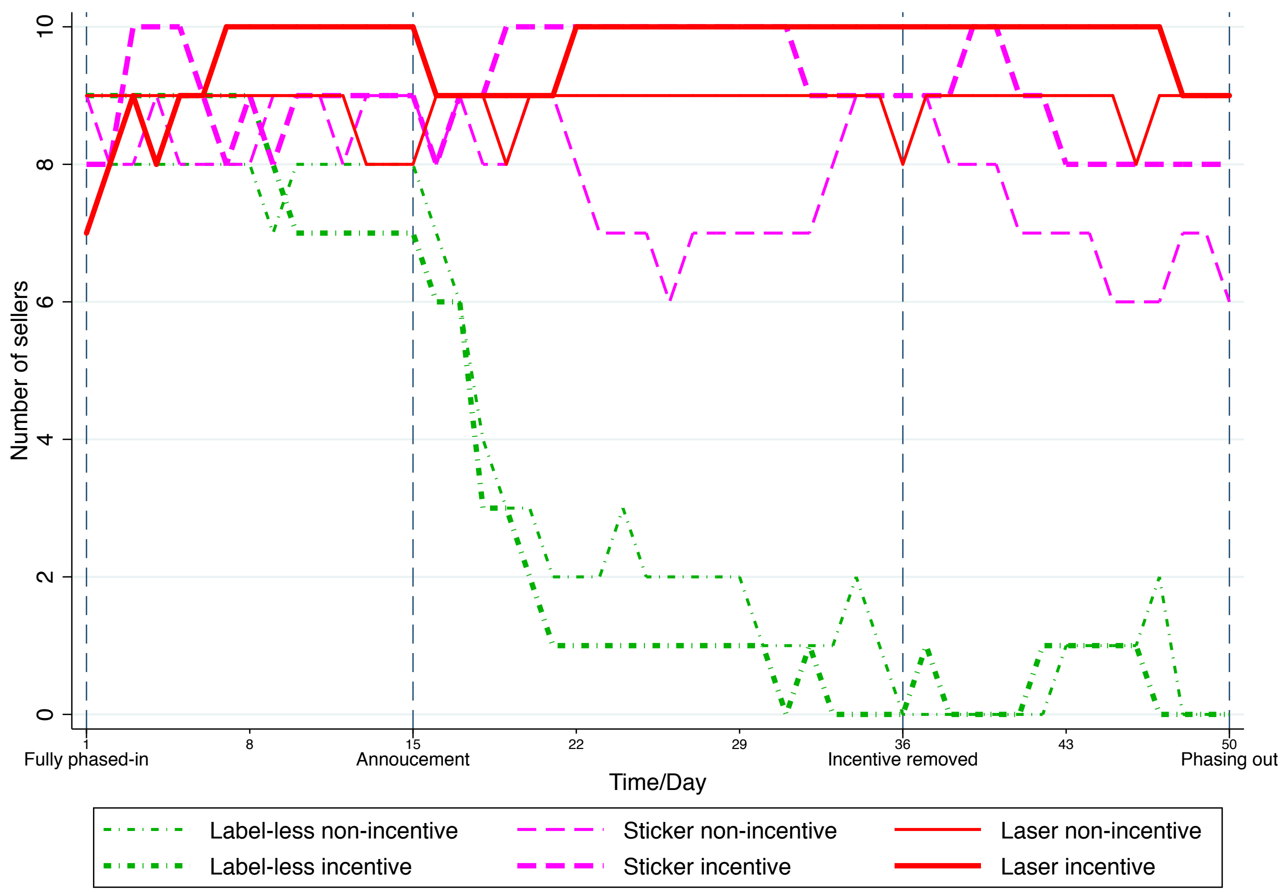
Figure 4: Timeline of the Study



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

1. A consumer pilot survey was conducted in December 2013 to elicit consumers' perceptions of different branding technologies.
2. Expressions of interests and baseline surveys were conducted in July 2014.
3. The market intervention was rolled in from July 13 to 19, 2014. The intervention was kicked off with the label-less group on July 13 and 14, followed by the sticker group on July 16 and 17, and finally the laser group on July 18 and 19th. July 19 is defined to be day 1 of the full-market intervention.
4. Quality differentiation was mandatory for the first 2 weeks, from July 19 to August 3. An announcement was made to all sellers on August 3 that they were free to differentiate or not afterwards.
5. On August 23, 35 days (6 weeks) into the intervention, the incentive (for the incentive groups) was lifted.
6. September 6 is the last day of the full-market intervention. An endline survey was conducted at surveyors' final visits to sellers' stores. Most of data analysis focuses on the period from July 19 (day 1) to September 6 (day 50).
7. The market intervention was gradually phased out from September 6 to September 12, 2014.
8. A short follow-up survey was conducted from September 14 to 20, 2014, and another one was conducted a year later, in July 2015.

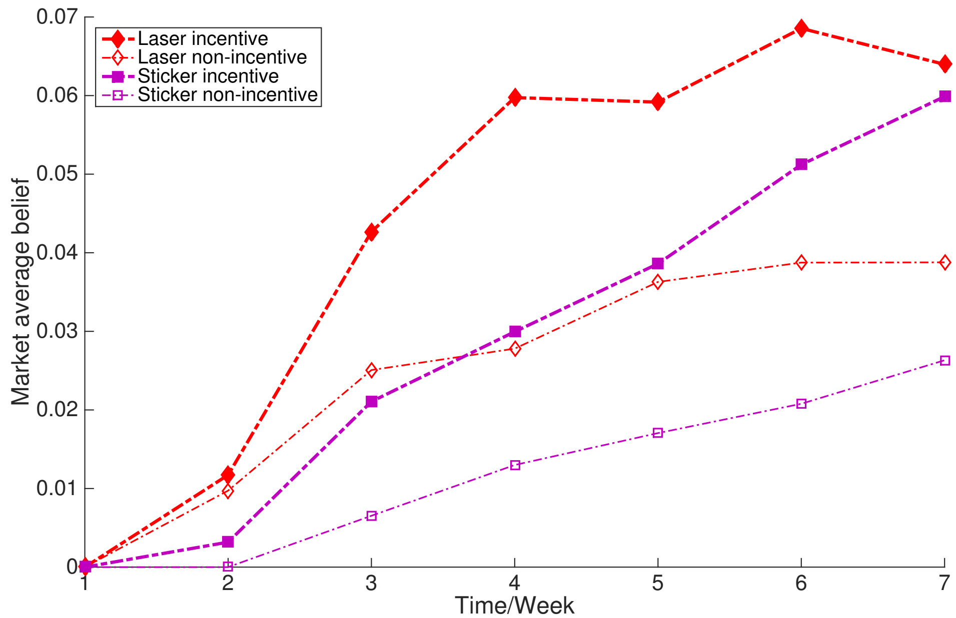
Figure 5: Quality Differentiation at Sale



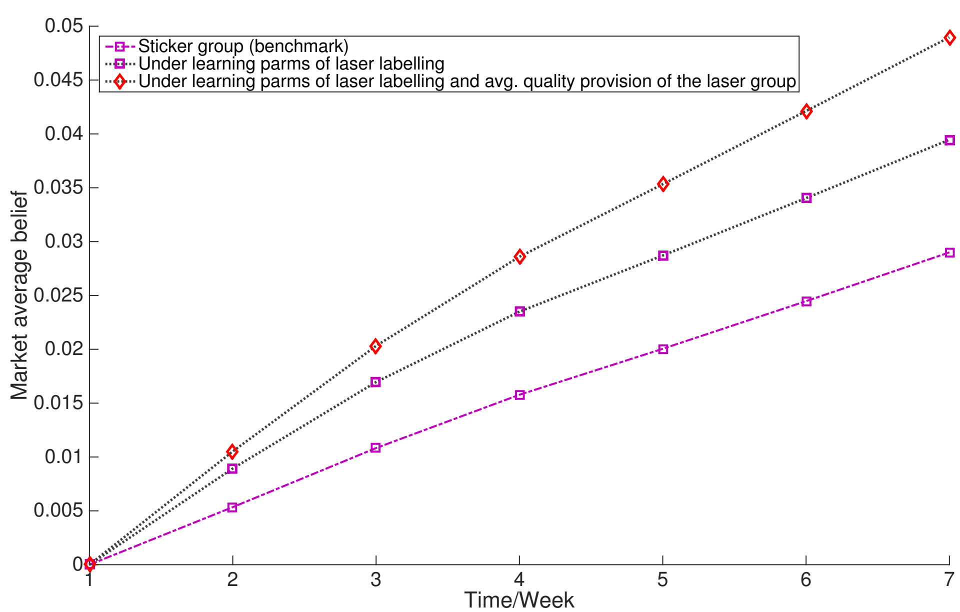
Note: This figure plots the number of sellers who differentiated quality at sale in each treatment group over time. The time axis runs from July 19 (day 1) to September 6 (day 50), 2014, corresponding to the period of the fully phased-in market intervention. The panel is not fully balanced because not all sellers operated their businesses on all days.

Figure 6: Beliefs Evolution

Panel A. Average Beliefs Evolution by Treatment Group

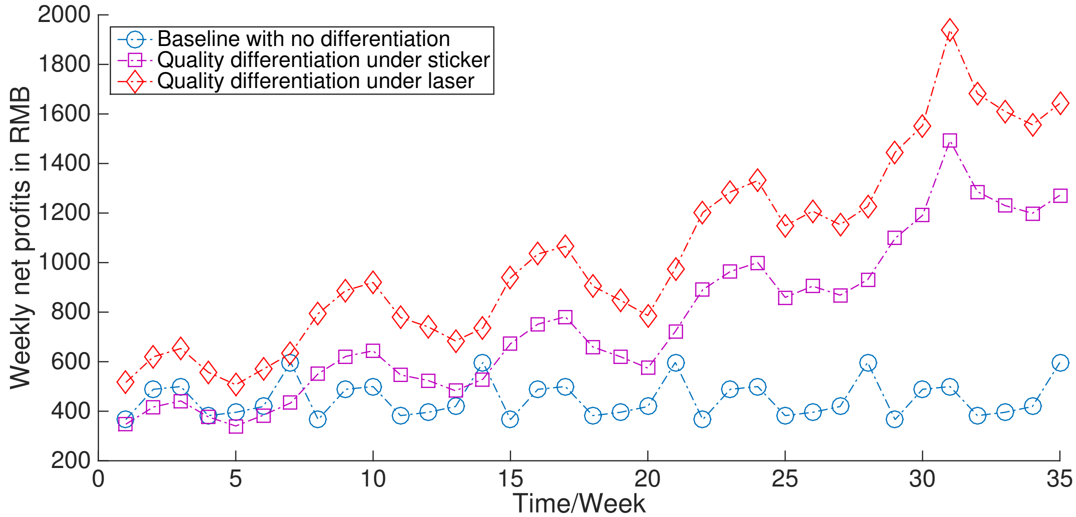


Panel B. Counterfactual Beliefs Evolution



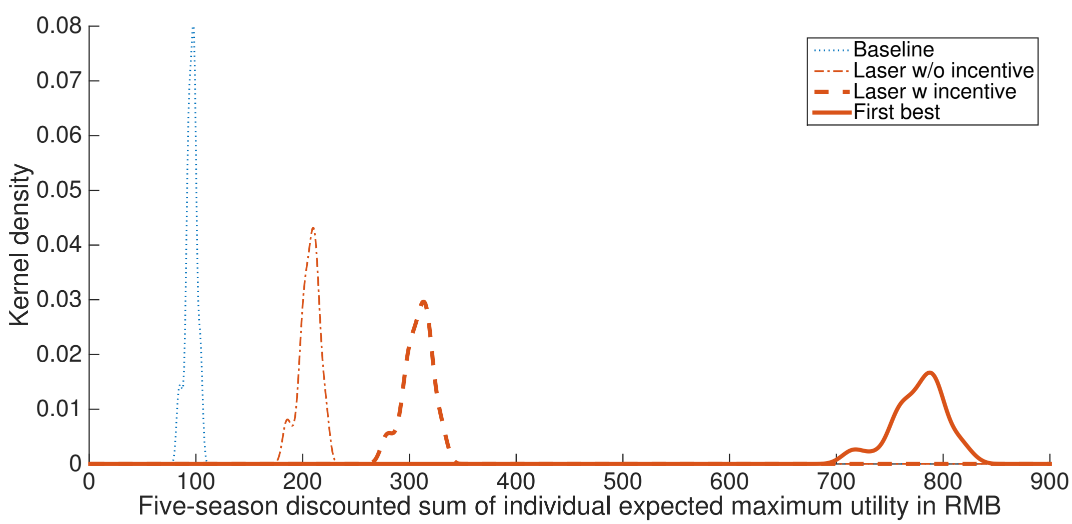
Note: This figure plots the average beliefs evolution about the quality of the premium pile. Panel A plots the market average beliefs calculated using the estimated prior beliefs (see Table 9) and the actual experience realizations for households in each treatment group. Panel B simulates the counterfactual beliefs evolution for the sample of households in the sticker group under three different scenarios: (1) under sticker group's average empirical quality (measured in terms of the empirical satisfaction rate for sticker-labeled watermelons); (2) the same quality as in (1) but replacing the prior beliefs with that under laser; (3) replacing both the prior beliefs and the average empirical quality with that for the laser group. The simulation procedure is discussed in Appendix D.3.

Figure 7: Net Profits Evolution



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

Figure 8: Distributions of Consumer Surplus



Note: This figure plots steady state distributions of consumer surplus for households in the hypothetical *average market* under the following four scenarios: (1) baseline with no quality differentiation; (2) quality differentiation under laser without incentive; (2) quality differentiation under laser with incentive; (4) first-best outcome. Consumer surplus is calculated as the five-season discounted sum of individual expected maximum utility in RMB (scaled by the price coefficients). Details for constructing the hypothetical market are explained in Section 6.2 and the formula for computing consumer surplus is shown in Section 7.

Table 1: Baseline Summary Statistics

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

Note: This table provides the summary statistics for sample characteristics of communities, sellers and households measured in the baseline surveys. In total, 60 sellers in 60 communities (markets) and 675 households were recruited for this study. Variation in the number of observations are due to missing responses in the baseline surveys. The measure for household's willingness to pay for quality is explained under the footnote of Figure 2.

Table 2: Purchasing Dynamics under Different Branding Technologies

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

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

Table 3: Quality Provision by Treatment Group

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

	A. By branding treatments (sticker and laser)				B. By incentive treatment (during incentive)			
	Non-incentive		Incentive		Laser		Sticker	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Laser	0.711***	0.619**	0.282*	0.309**				
	(0.222)	(0.266)	(0.136)	(0.128)				
Incentive					0.496*	0.563**	1.033***	1.006***
					(0.246)	(0.266)	(0.176)	(0.176)
Observations	238	238	230	230	197	197	194	194
Baseline Controls		✓		✓		✓		✓
Time Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Omitted group mean	9.738		10.654		10.451		9.738	
Std. dev	(1.104)		(0.886)		(1.04)		(1.104)	

Note: This table examines quality provision by treatment group. Quality is measured in sweetness. Each observation is at the seller-check level. The key explanatory variables are the group dummies. The mean and standard deviation for the omitted group are shown in the bottom two rows. Panel A examines the heterogeneity across different branding groups. Columns 1 and 2 restrict the sample to the non-incentive groups only. Columns 3 and 4 restrict to the incentive groups. Panel B examines the heterogeneity for sellers with and without the incentive. Since sellers in the label-less group reverted back to non-differentiation after the mandatory period, the sample for this analysis includes only sellers in the sticker and laser groups. The time period is from week 1 to week 6, before the incentive was lifted. Columns 5 and 6 look within the laser group. Columns 7 and 8 look within the sticker group. All regressions control for check fixed effects. The even columns control for 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. Standard errors are clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 4: Quality Differentiation Behavior

Sample: sticker and laser non-incentive groups

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

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

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

Sample: non-incentive groups

	Ln(Sales Profits)		Premium Markup		Premium Quantity		Normal Markup		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	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Label-less Mean	4.284		0.055		56.313		0.011		180.475		236.788	
Std. dev.	(0.687)		(0.091)		(136.508)		(0.084)		(124.07)		(156.597)	

Note: This table examines sales profits, price and quantity for sellers in the non-incentive groups. Each observation is at the seller-day level. Sticker and laser are group dummies, and the omitted group is the label-less group, the mean and standard deviation for which are shown in the last two rows. Markup is defined to be the difference between the unit price (RMB/Jin) charged by the seller and the market average. Quantity is measured in Jin and profits are measured in RMB. If a seller stops to differentiate quality at sale, the unit price of the premium pile is defined to be the same as that of the normal pile, and the sales quantity of the premium pile is coded as 0. The even columns control for the following set of seller and community baseline characteristics: number of competitors in the local market, average housing price, and distance to the nearest supermarket. All regressions control for day fixed effects. Standard errors are clustered at the seller level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

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

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

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

Table 7: Household Endline Perceptions

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

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

Note: This table examines the endline willingness to pay for quality for households in different markets. The dependent variables are the maximum self-reported prices (in RMB/Jin) households are willing to pay for watermelons under different branding technologies (see footnote of Figure 2 for details on eliciting the willingness to pay). Coefficients on the branding treatments and the interactions between the branding treatments and the incentive treatment are reported in this table. The omitted group is the label-less non-incentive group. In addition, the even columns control for a set of household baseline characteristics, including household size, percentage of elderly, monthly income, average number of watermelons consumed per week reported in the baseline survey, and the baseline self-reported willingness to pay for quality (measured in RMB/Jin). Standard errors are clustered at the market level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 8: Effects of Removing the Incentive on Quality Provision

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

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

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

Table 9: Simulated Maximum Likelihood Estimation Results of Consumer Learning Models

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

Note: This table shows the simulated maximum likelihood estimation results of the consumer learning models. a_0 and b_0 are constrained to be non-negative. Details for the estimation procedures are explained in Appendix D.1. Column 1 shows the estimates for the baseline model. Column 2 considers direct utility of laser label. Column 3 incorporates reputation spillover across a seller's two piles. Column 4 considers information diffusion. The log-likelihood ratio statistics for testing the baseline model against these alternative models are presented in the last row. Estimates for the market and time fixed effects are abbreviated from this table and are reported in Appendix Table 8. Standard errors shown in parentheses are calculated using the outer product of gradients (OPG) estimate for the asymptotic covariance matrix (see Appendix D.3 for details).

Table 10: Simulated Market Outcomes

Structural parameters

Market size : 4.5×194 households (to match initial sales quantity)

$\delta = 0.98$, $c = 0.64$

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

Note: This table simulates market outcomes for the hypothetical *average market* using the estimated dynamic demand system and the estimated supply-side parameters. Details for constructing the hypothetical market are explained in Section 6.2. Column 1 simulates the market outcomes under the average empirical policies of the laser non-incentive group and column 2 does that for the laser incentive group. Column 3 performs a counterfactual exercise by replacing the learning parameters (including $a_0, b_0, \Delta q$) under laser with those under sticker (see Table 9). Column 4 simulates the outcomes for the baseline case with no quality differentiation. Details for the simulation procedures are explained in Appendix D.3.

Table 11: Welfare Effects of Information Frictions and Fragmented Markets

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

Note: This table examines the welfare effects of information frictions and market competition. The top panel solves for the optimal policies under each counterfactual scenario. Quality is the probability of being good and markup is the difference between the prices of the premium and the normal pile, measured in RMB/Jin. The middle and bottom panel calculate the 5-season discounted sum of surpluses (in RMB) under the corresponding policies for the same hypothetical *average market* as that for Table 10 (see in Section 6.2 for details on constructing the hypothetical market). Details for calculating the consumer and producer surpluses are discussed in Section 7.