Essays on the Industrial Organization of Retail Markets

by

Tiancheng Chen

Department of Economics
Duke University

Date: ______________________________

Approved:

____

Allan Collard-Wexler, Advisor

____

Yi (Daniel) Xu, Co-Advisor

____

James Roberts

____

Carl Mela

____

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2021

ABSTRACT

Essays on the Industrial Organization of Retail Markets

by

Tiancheng Chen

Department of Economics
Duke University

Date: ______________________
Approved:

____

Allan Collard-Wexler, Advisor

____
Yi (Daniel) Xu, Co-Advisor

____
James Roberts

____
Carl Mela

____

An abstract of a dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2021

Abstract

I study the industrial organization of both online and offline retail markets, trying to uncover the impact of recent trends in this industry, namely store brand and big data. Chapter 2 gives a detailed discussion about the data collection and analysis process in both online and offline retail markets. I first describe two offline retailer scanner datasets provided by Nielsen, the Homescan dataset and the Retail Scan dataset. The former documents household shopping trips, and the latter provides detailed information of how many of each product are sold in every retailer. Then I move on to describe the datasets collected at a big Chinese e-commerce platform. I will talk about what the datasets look like, how the datasets are collected and organized, and finally how they are analyzed.

Chapter 3 studies the welfare consequences of store brands in the offline retail market. I model the vertical relationship between retailers and manufacturers with a Nash-in-Nash bargaining model. I show that stronger preference for store brand has an ambiguous effect on the non-store brand prices, and that this effect is nonlinear: it is more negative when store brand has a smaller market share. Using Nelson Homescan data. I find a 1% increase in store brand's market share leads to a more than 0.5% decrease in non-store brand prices. This shows that in reality, the main impact of store brand is to help retailers gain better bargaining positions vis-à-vis suppliers. Furthermore, I also find the negative effect is larger in magnitude when the market share is smaller.

Chapter 4 studies the impact of market intelligence data on online retailer performance, product choice, and market outcomes. To do so, I exploit a unique setting in which an e-commerce platform provides its sellers with a market intelligence tool called Market Insight. First I find Market Insight helps online sellers choose better products and increases their sales. Second I show the current design of Market Insight benefits consumers and the platform, though providing "too much" information through Market Insight could be harmful. Finally I solve for the platform-optimal design of Market Insight and show that the total sales revenue on the platform would increase 8% under this design, and consumer welfare will increase 0.8%. I also compare the platform-optimal design to the socially optimal design. They are very close to each other, showing in this case, the platform acts like a benevolent social planner.

Acknowledgements

I would like to thank my advisors, Allan Collard-Wexler and Daniel Xu for their for their invaluable advising, continuous support, and above all, their patience and tolerance with me throughout my years at Duke. They are great mentors and excellent friends. I feel very lucky to be their student. They make my PhD experience both difficult and rewarding, and the most precious memory to hold on to. I am grateful to my committee members, James Roberts and Carl Mela, for their great advice. I have appreciated and enjoyed all our discussions and conversations. I am also grateful to Adam Rosen, Matt Masten, Attila Ambrus, and Todd Sarver. Discussions with them have always been most intellectually rewarding.

I would like to thank my partner, Yilin Jiang, for her love, friendship, and company. She made every day of the last seven years happier and tenderer for me. I am also grateful to my best friends, Yanyou Chen and Tashi Sun. Talking with them has always been the best way to waste time.

I am grateful to my parents, Jianhua Chen and Lixin Yao. They have always been supportive even though they are on the other side of the globe. Their love has been the true source of my power to keep going.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Retail industry is experiencing a lot of changes compared with a decade ago. First and foremost, e-commerce is profoundly changing our shopping experiences. In the US, when a consumer wants a chair, very likely she will simply find it on Amazon or other websites and get it delivered in a few days. In fact, one can buy almost anything online, ranging from small appliances to meats and vegetables to cars. In China, online shopping is developing even faster thanks to a few dominant tech giants. On Alibaba, for example, even houses could be auctioned and bought.

Secondly, the store brands are playing a bigger role. For the offline retailers such as Walmart, store brands serve as a tool both to attract consumers and to bargain with upstream suppliers. Online retailers are also keen to offer their own store brands, taking advantage of their ability to collect "big data" of consumption behaviors and to control the display of product offerings.

A third, maybe more fundamental trend is the progression in data storage and collection technology and data science. Both online and offline retailers try to take advantage of this trend, and to make more data-driven decisions.

While there're so many changes in this very old industry, our understanding on the impact of these trends are lacking. In this thesis I try to provide three distinct yet inter-related analyses in order to give a clearer understanding of 1) how is data collected and analyzed in both online and offline retail markets, 2) what is the impact of store brands on consumers, and 3) how do online sellers use market intelligence data and how should an e-commerce platform design its information policy?

Chapter 2 gives a detailed discussion about the data collection and analysis process in both online and offline retail markets. While the scanning data at offline retailers are more familiar to economists, what is the "big data" collected by e-commerce platforms, how are they collected, and furthermore how do data scientists analyze these data are less known. In this chapter I first describe two scanner datasets provided by Nielsen, the Homescan dataset and the Retail Scan dataset. The former documents household shopping trips, and the latter provides detailed information of how many of each product are sold in every retailer. Then I move on to describe the datasets collected at a big Chinese e-commerce platform. I will talk about what the datasets look like, how the datasets are collected and organized, and finally how they are analyzed.

Chapter 3 analyzes the welfare consequences of higher private label market share. I characterize two effects: the direct effect is that store brands benefit consumers since they have lower price. The indirect effect is that store brand affects consumer welfare through affecting the non-store brand prices. To characterize this indirect effect, I build a Nash-in-Nash bargaining model taking into account the vertical relationship in retail industry, namely manufacturers of national brand sell to a retailer who combines a lot of different products and then sell to final consumers. In particular, the wholesale price at which manufacturers sell to retailers is determined through bargaining. In addition to national brands, retailers sell their own store brands, which are typically sold at lower prices but generates a higher margin, as store brands are usually acquired at marginal cost (Steiner, 2004 [47]). Therefore in the model, the store brands not only compete directly with national brands, they also help retailers to gain better bargaining positions.

Using this model I characterize two distinct and countervailing effects of higher preference for store brand. The first is pricing effect: a higher store brand preference leads to a higher national brand price since the retailer wants to raise store brand price and internalizes business stealing effects. The second is bargaining effect: a higher store brand preference generates a stronger bargaining position of the retailer, thus lower wholesale price. The key insight is that the indirect effect is more negative when the store brand has a smaller market share, as in this case, the bargaining effect dominates the pricing effect.

Using Nelson Homescan data, I find that first, on average 1% increase in absolute store brand market share leads to more than 0.5% decrease in non-store brand prices. Furthermore, splitting sample based on store brand market share, I also show that the negative effect is larger in magnitude when the market share is smaller.

Chapter 4 studies the impact of market intelligence data on online seller performance, product choice, and market outcomes. To do so, I exploit a unique setting in which an e-commerce platform provides its sellers with a market intelligence tool called Market Insight. The main findings are four-fold. First, Market Insight helps online sellers choose better products and increases their sales. An exogenous policy change of Market Insight that gave some sellers more information led to 40% higher revenue among treated sellers, raised their rate of introducing new products by 20%, and, notably, resulted in them selling more profitable products. Second, Market Insight is valuable to consumers and to the platform; it increases consumer welfare and total sales by 0.8% and 8%, respectively. Third, platform sales and consumer welfare are not monotonically increasing with the amount of information provided; when sellers are more aware of market trends, they drop unprofitable products, which could reduce product variety and harm consumers and the platform. Finally, I solve the optimal design of Market Insight that maximizes platform sales, and then compare it to the socially optimal design. They are very close to each other, showing in this case, the platform acts like a benevolent social planner.

Chapter 2

Datasets of Online and Offline Retailers

2.1 Introduction

With the progress in data technology and data science, firms more and more emphasize on the “data-drive” decision-making process. One salient example is retailers. Offline retailers have a long tradition to use their own scanner data to manage their product selection, inventory, pricing, as well as store-brand offerings. Data vendors including Nielsen provide not only scanner data from all major offline retailers, but also data on household purchasing trips. Combined, retailers get more information about competitors and consumption patterns.

Online retailers could get even more information about consumers. They observe not only the product bundles each consumer ends up buying, but also the consumer’s whole chain of actions. Assume a buyer wants to buy a lipstick, then the e-commerce platform (e.g., Amazon) sees the keyword used by the consumer, how many different items are browsed, which ones she clicks on, how long she spent on each different page, whether she adds an item to her shopping cart, and after how long she buys it. The platform also sees whether she returns the item or requests refund. This combined with the whole shopping history of each buyer gives the e-commerce platform a very detailed depiction of the shopping patterns of each buyer.

2.2 Nielsen Datasets for Offline Retailers

In section 3 I study how store brand affects non-store brand prices. For that I use two datasets from Nielsen. The first is the Nielsen Homescan dataset. It records shopping trips of 40,000-60,000 panelists across US from 2004 to 2015. A shopping trip contains information of the store of the trip, the products (at barcode level) brought, the total spending, and the date of the trip. It also records anonymously the retailer of the store, i.e., if it's Target or Walmart. There's also detailed information of the panelist, such as household size and income.

In this dataset, a product is defined at barcode (UPC) level. In Homescan data, products are organized into departments, product groups, product modules, and UPC codes. Departments, product groups and product modules are all Nielsen defined codes, while UPC codes are defined by manufacturers[1]. The hierarchy is as follows

[1]with the exception of Magnet products, which are grouped into Nielsen-generated UPCs

- Department (10 Departments)
  - Product Group (125 Product Groups)
    - Product Module (1,075 Product Modules)
      - UPC (3.2 million UPC Codes)[2]

The Homescan data will be used to construct store brand market share at product module level. The market share is calculated as total units of store brand sold (of a particular product module) divided by total units sold of that product module, within a retailer within a time period. Important for this construction is the variable *projection factor* reported in the Homescan data. The household specific projection factors are sampling weights of each household in Nielsen's survey. The sum of the weights is the total household population (i.e. number of households) in the U.S.

One shortcoming of the dataset is that it doesn't have precise information on location. For panelists, it records which DMA[3] they live in. For a store, it records the first three digits of store zip code. But since in this practice I aggregate sales at retailer level, and focus only on mass merchandisers who typically operate nation-wide, location information is not important as of now. However, it conceivably restricts future steps such as estimating travel cost.

Another problem is that if sales of a product is small, then it shows up even more rarely in the Homescan dataset. This is particularly an issue with new product entry and old product exit; for both of these kind of products tend to have small sales. The Nielsen Retail Measurement Services (RMS) scanner dataset could be really helpful for this problem, as it reflects the universe of purchases as opposed to the purchases of a sample of households. The dataset consists of more than 100 billion unique observations at the week-store-UPC level. Each individual store reports weekly prices and quantities sold for every UPC code that had any sales volume during the week. The data contains approximately

[2]Note that 3.2 million UPC codes are contained in the combined consumer panel and retail scanner data files. Only 1.5 million UPC codes are present in the consumer panel files only.

[3]DMA is the market defined by Nielsen. In 2015, there're totally 218 DMAs in the US.

40,000 distinct stores from 90 retail chains across 371 MSAs and 2,500 counties between 2006 and 2014. Over the sample period the total sales across all retail establishments are worth approximately 2 trillion and represent 53% of all sales in grocery stores, 32% in mass merchandisers (Argente et al, 2017).

Though RMS has more precise information, it only reports store sales for a subset of stores; thus making it impossible to infer retailer level sales. So I will use this dataset later when I need to define product entry and exit, but for the current practice, only Homescan data is used.

2.3 Datasets of an E-Commerce Platform

In section 4 I study how online retailers make use of market trend data. For this study I went to Alibaba, the largest e-commerce platform in China. Like Amazon, Alibaba collects a huge amount of data on its buyers and sellers. Several datasets document the detailed purchasing behavior of each buyer. One dataset documents the viewing pattern: one observation corresponds to one consumer viewing one item at one time. With each observation there is also a keyword input by the consumer for that particular search, as well as the seller identity who sells the item. Another dataset documents the clickstream; one observation corresponds to one consumer clicking on one item at one time. The clickstream dataset could be linked back to the viewing data. There're also datasets about consumer behavior after she clicks on an item: does she add it to her shopping cart? Does she buy it directly? There're other datasets that document consumer behavior with her shopping cart, such as when an item is added, and when a transaction takes place.

On the seller side, the platform also keeps a detailed documentation of their behavior. For example, the platform would know on each day, what products each seller sells. The platform also keeps track of the seller reputation (e.g., reviews), since one big function of the platform is to tell consumers who are excellent sellers. Additionally, the platform keeps track of the seller (on-the-platform) advertising spending and shipping service. Generally speaking, the platform finds a way to collect all the data about actions taken on it, but has little information otherwise. One important piece of data that this platform is missing (and probably for many other e-commerce platforms) is the supply side of the sellers. For example the platform doesn't know who are the suppliers of the online sellers, and what are the input prices. Another important piece of missing data is competitors' behavior. For example, the platform doesn't know how many consumers go to competing platforms, how many sellers there are, and how much sales conducted. One might imagine that sales information at other platforms could be scraped but this is actually pretty hard; each e-commerce platform highly values their own data and tries very hard to ban data scraping from third-party users (at least in China).

From the raw data, numerous datasets are created. For example, from the consumer purchasing behavior one could generate an aggregate sales dataset of each seller. Each dataset is created and maintained by an individual, and anyone who uses the data could reach directly to the data creator to ask the meaning of each variable as well as the underlying code to figure out what are included and what are left out. The data owner also determines who gets to see the dataset and who doesn't, following certain regulations that aim to protect the privacy of the buyers and sellers on the platform.

Finally, to conduct analysis based on these data, a data scientist in many cases will need to construct his own dataset comprised of certain variables and aggregated to certain level (seller level, monthly level, category level, etc). This is done through SQL on the cloud computer owned by the platform. If more complicated analysis (say, OLS regression) is required, then the data scientist could use languages installed on the could server.

For the study in Chapter 4, I constructed a dataset of Taobao lipstick sellers, consisting of monthly data from January 2019 through October 2019. For each seller that offers lipsticks at any point during the sample period (regardless of sales results), we observe items carried, units sold, and sales prices. More importantly, we also observe their subscription to Market Insight within each month (basic, pro, or nonsubscriber). I also observe a bunch of seller characteristics including Gold seller status, Global seller status, number of stars, and number of fans.

Gold status is a certificate of excellence for Taobao sellers.[4] Global status indicates a seller who purchases all products overseas and sells to consumers in China. Number of stars represents a seller's total number of transactions.[5] Fans represents the number of consumers who have signed up to be a fan of the seller; fans can find the seller more easily and get notifications of the seller's new products.

[4]To qualify, a seller must have relatively high ratings, have a low complaint rate, offer a standard return policy, and, most importantly, pass a minimum sales requirement. The standard of sales varies by category; for cosmetics, a Gold seller has to have made at least 400 transactions with total sales of $\geq 3,568$ USD during the previous six months.

[5]After each transaction, a seller gets positive or negative feedback from the buyer; if the buyer chooses not to submit feedback, the feedback becomes automatically positive. Star level is derived from the total amount of positive feedback minus negative feedback, and therefore represents roughly the total number of transactions.

Chapter 3

The Effect of Store Brand Market Share on Non-store Brand Prices

3.1 Introduction

Private labels[1] account for a large share of retail industry revenue: in 2012, they generated $98 billion in U.S. food and grocery sales, or 17.1% of total revenues (Schultz, 2012 [**Schu/02**]; Nielsen Global Survey, 2012). Not only that private labels market share is significant, but it has been increasing for a long time. Its fast growth back in 1987-2000 has been well documented (see Hoch et al, 2002 [23], and Steiner, 2004 [47] for example). But even now, though slowly, the market share is still growing. According to Nielson, from 2009 to 2016, the market share steadily grows from 16.1% to 18.4%.

It is therefore important to understand welfare consequences of higher private label market share. The direct effect is that store brands, through lower prices compared with national brands, generate benefits to consumers. The indirect effect, the effect of a higher store brand market share on non-store brand prices, is also important to welfare analysis. However, this indirect effect is difficult to characterize, mainly due to the vertical relationship in retail industry: manufacturers of national brand don't sell directly to consumers; they sell to a retailer who combines a lot of different products and then sell to final consumers. The wholesale price at which manufacturers sell to retailers is determined through bargaining. In addition to national brands, retailers sell their own store brands, which are typically sold at lower prices but generates a higher margin, as store brands are usually acquired at marginal cost (Steiner, 2004 [47]). Therefore the store brands not only compete directly with national brands, they also serve as a tool of retailers to gain better bargaining

[1] I use store brand and private label interchangeably in this paper.

positions.

To characterize this indirect effect, I build a simple model that captures this vertical relationship. In the model, a retailer sells two products, one store brand, and one national brand whose wholesale price is determined through bargaining with its manufacturer. Using this model I characterize two distinct and countervailing effects of higher preference for store brand. The first is pricing effect: a higher store brand preference leads to a higher national brand price since the retailer wants to raise store brand price and internalizes business stealing effects. The second is bargaining effect: a higher store brand preference generates a stronger bargaining position of the retailer, thus lower wholesale price. The key insight from the model is a nonlinear effect that depends on the store brand market share: when it is small, then the bargaining effect dominates and vice versa. This means the marginal effect of store brand market share should be more negative when it is smaller.

I test this prediction using Nelson Homescan data. To get consistent estimates, I use the median of store brand market share within a retailer to instrument store brand market share in a particular product module. The logic is that tastes of and advertisements for products under the same store brand name are correlated, yet production costs are not. Using this IV, I find that first, on average 1% increase in absolute store brand market share leads to more than 0.5% decrease in non-store brand prices. This means the bargaining effect dominates the pricing effect. Furthermore, splitting sample based on store brand market share, I also get some evidence that the negative effect is larger in magnitude when the market share is smaller.

The rest of the chapter is organized as follows. Section 2 talks about related literatures. Section 3 lays out the model and discuss the pricing effect and bargaining effect in more details. Section 4 talks about the data. Section 5 presents estimation and results. Section 6 concludes.

3.2 Literature Review

This study is related to several strands of literature. The most related literature talks about the bargaining benefit of private labels. Using retail and wholesale data of four categories in the Dominick's retail chain, Pauwels and Srinivasan (2004) [37] finds that introduction of store brands leads to higher unit margin on national brands, indicating that the retailers strengthening bargaining position vis-a-vis national brand manufacturers. Chintagunta et al. (2002) [8] also use Dominick's data on the oat category to study entry of store brands. They also find that store brand introduction coincides with an increase in the retailer's margin for the national brand. Besides, they find consumer preference for national brands are relatively unaffected by the introduction of store brands. Meza and Sudhir (2010) [33] again uses Dominick's data, but on the breakfast cereal category. They estimate a structural demand model and compare its suggested wholesale price with true wholesale price from data, and find that after store brand entry, manufacturers charge a lower than the profit maximizing price, thus supporting the bargaining benefit story.

With the advances in Nash bargaining models, a few recent papers use Nash-in-Nash protocol to structurally model and estimate the bargaining process between retailer and supplier. Draganska et al (2010) [14] looks at coffee market in Germany. Using only sales data, they're able to back out bargaining power and wholesale prices, and find that in this market, bargaining power lies mainly with manufacturers. In a related study, Norton and Elberg (2016) [35] again looks at coffee market, but in Chile. In addition to sales, they also have access to wholesale data. They find that compared with small manufacturers, big suppliers have a better bargaining position and higher bargaining power. Some suppliers also have high bargaining power, and can obtain good bargaining position through its loyal customers. Ellickson et al (2017) [16] also looks at the coffee market, this time in the US. Exploiting the event of patent expiration of Keurig's 'K-Cup' technology, and subsequent entry of store brands, they find enhanced retailer bargaining position accounts for roughly 20% of the overall benefit of private label entry.

There's a broader literature on private labels. A classic topic is to identify the factors of private label success through cross category analysis. Hoch and Banerji (1993) [24] examines this question from the manufacturer and retailer perspectives, and finds that large category, high margins, fewer national manufacturers and advertising spending from those manufacturers contribute to higher private label shares. Raju et al (1995) [38] distinguishes the determinants of store brand introduction vs market share performance. While lower cross-price sensitivity among the national brands, higher cross-price sensitivity between the national brands and the store brand, and a higher base level of demand for store brand lead to more entry and higher market share of store brand, larger number of national brands leads to more entry but lower market share of store brands. Sinha and Batra (1999) [45] adds the consumer side by examining what role perceived risk[2] plays in success of store brands. They find that private label purchase is higher in categories with lower perceived risk. Later work by Hansen et al. (2006) [22] looks at ten food and nonfood product categories and find that the tendency to buy store brand is a household attribute rather than category specific.

Another topic is the positioning of store brands. Sayman et al. (2002) [43] studies the positioning of store brands and identifies conditions under which the optimal positioning strategy is to be as close as possible to the stronger national brand. Morton and Zettelmeyer (2004) [34] shows that positioning private labels close to national brands enhances retailer bargaining position by lowering the value added by the national brand.

The Nash-in-Nash framework used in my application is used by a lot of papers. Worth noting are Horn and Wolinsky (1988) [26], who give the theoretical development of the Nash-in-Nash solution, Crawford and Yurukoglu (2012) [12], who apply the framework to model bargaining between content providers and distributers, Grennan (2013) [20], who apply the framework to model bargaining between hospitals and stent manufacturers, Collard-Wexler et al. (2014) [9], who recent provide a non-cooperative foundation for "Nash-in-Nash" bargaining that extends the Rubinstein (1982) [41] alternating offers model to multiple upstream and downstream firms.

[2] They use several measures of perceived risk, such as quality variability

3.3 A Model of Retailer Bargaining

As discussed before, to learn welfare consequences of store brands, it's important to understand how presence of store brands affect non-store brand prices. In this section I build a very simple model to study how an increase in preference of store brands affects non-store brand price. There are multiple reasons that could explain higher store brand market share, e.g., lower (higher) cost of store brand (non-store brand), or higher (lower) preference of store brand (non-store brand). Here I focus on higher preference of store brand because I think it is the main reason for higher store brand market share as of today compared with 30 years ago.

To understand how change in preference in store brands will affect non-store brand prices, it's key to understand the role of store brands in this setting. On one hand, a retailer earns profits from both store brands and non-store brands. The difference between a retailer in this setting and a multi-product firm that maximizes joint profit from all his products is that here a retailer also needs to bargain over the wholesale price for non-store brands. A stronger store brand also gives the retailer a stronger bargaining position, in the sense that in case he didn't reach agreement with a non-store brand manufacturer to carry its product, he could still capture a relatively large market share with its own store brand.

Therefore an increase in preference for store brands has two distinct and countervailing effects on non-store brand prices. On one hand, now the retailer has an incentive to raise its store brand price to take advantage of the higher preference. To do this, he also wants to raise the non-store brand price as keeping it low, the non-store brand would steal market share from the store brand. I.e., the retailer has an incentive now to raise non-store brand price since he internalizes the business stealing effect. I call it the pricing effect. On the other hand, as mentioned above (and will be characterized in more details below), an increase in preference for store brands forces down the wholesale price. I call this the bargaining effect.

Which effect dominates? Intuitively it depends on relative market shares of the store brand vs the non-store brand: if the store brand already commands a market share close to one, then first, the retailer cares more about raising the store brand price, and second, the wholesale price would be already very close to marginal cost and could hardly be even lowered. Therefore, the pricing effect should dominate in this case; the non-store brand price should go up. However, if the store brand has a small market share but a non-store brand dominates the market, then the reverse should be true.

In subsection 3.1 I first lay out a simple model with only one retailer selling two products, one store brand and one non-store brand. Then in subsection 3.2 and 3.3, I characterize the pricing effect and the bargaining effect separately. Section 3.4 combines these two effects and show the intuition talked above is actually the case. Finally in section 3.5 I include some discussions of the model.

3.3.1 Model Setup

I assume there is one retailer, R, that sells two products.[3] Product 1 is the retailer's store brand, which is acquired by R at manufacturing cost c_1.[4] Product 2, on the other hand, is made by manufacturer M at cost c_2, and sold to retailer at wholesale price w. The wholesale price is determined through a bargaining process. I further impose a key assumption that the wholesale price and retail price are set simultaneously. I.e., when bargaining the wholesale price, the manufacturer will treat the retail price as fixed.[5]

Demand for product j is given by a simple logit model

$$u_{ij} = \delta_j - \alpha p_j + \epsilon_{ij}, \forall j \in \{1, 2\} \tag{3.1}$$

[3]this simple model abstracts away from a lot of things, including retailer competition, non-store brand competition, etc. See section 3.5 for more detailed discussion

[4]most of the theoretical work and all of the empirical work except Chen et al (2010) [7] assumes that PL products are procured in perfectly competitive fashion at marginal cost. See Chen et al (2010) [7]

[5]One may argue that a more natural assumption is that retail prices are set conditional on all wholesale prices as in a manufacturer Stackelberg game. However, some theory papers adopt the simultaneous move setup, e.g., Iyer and Villas-Boas (2003) [29] and Dukes et al. (2006) [15] and argue that it is consistent with retail prices unobservability: retailers may exert effort to promote the item and thus charge a higher retail price; the retail price is not closely monitored by manufacturer

and the utility of not purchasing is

$$u_{i0} = 0 + \epsilon_{i0}$$

where ϵ is iid distributed EV1. Consumer i chooses $j = 0, 1, 2$ to maximize his utility. Then from the property of extreme value distribution, we know product j's market share is given by

$$s_j = \frac{exp(\delta_j - \alpha p_j)}{1 + \Sigma_{k=1,2} exp(\delta_k - \alpha p_k)}, \forall j \in \{1, 2\}$$

and

$$s_0 = \frac{1}{1 + \Sigma_{k=1,2} exp(\delta_k - \alpha p_k)}$$

The retailer sets (p_1, p_2) to maximize retailer profit

$$\Pi_r = (p_1 - c_1)s_1 + (p_2 - w)s_2 \tag{3.2}$$

given an equilibrium belief of w. Note I assume that the retailer doesn't know w when he sets the retail prices.

The retailer R and the manufacturer M bargain over the wholesale price w, which, according to Nash bargaining literature, should solve the Nash product:

$$np = (\Pi_r - D_r)^{b_r} (\Pi_m - D_m)^{1-b_r} \tag{3.3}$$

Here b_r is the retailer bargaining power, and $1 - b_r$ is the manufacturer bargaining power. For simplicity I assume $b_r = \frac{1}{2}$ to derive results below.

Π_r is the retailer profit if agreement is reached, D_r is disagreement value to retailer. Since changing w here will not affect retail prices (p_1, p_2), Π_r here is retailer profit under (p_1, p_2) and w, and D_r will depend on p_1 only, since now that agreement cannot be reached, retailer R doesn't carry product 2, and p_1 cannot change. Π_m is the manufacturer profit if agreement is reached, D_m is disagreement value to manufacturer, also fixing (p_1, p_2). In particular, since manufacturer M has only one product, $D_m = 0$

3.3.2 Pricing Effect

As discussed before, a higher preference for store brands (in this model, an increase in δ_1) affects non-store brand price (p_2) through two ways. The first way, the pricing effect, refers to that R has an incentive to raise p_1 to take advantage of a higher δ_1. To do this, he also wants to raise p_2. In particular, I want to solve $\frac{\partial p_2}{\partial \delta_1}$. Note that in this practice I want to separate the pricing effect from bargaining effect, so I will fix w for now.

The idea will be to first write out the first order condition of the retail pricing problem. This gives an implicit function of $(p_2, w, \delta_1, \delta_2)$. w is there because I hold it fixed. Then I use the implicit function theorem (IFT) to derive $\frac{p_2}{\delta_1}$.

Step 1: optimal retail price

Recall the retailer sets (p_1, p_2) to maximize equation [2]. This gives the FOC:

$$\begin{cases} s_1 - (p_1 - c_1)\frac{\partial s_1}{\partial p_1} + (p_2 - w)\frac{\partial s_2}{\partial p_1} = 0 \\ s_2 - (p_2 - w)\frac{\partial s_2}{\partial p_2} + (p_1 - c_1)\frac{\partial s_1}{\partial p_2} = 0 \end{cases}$$

After some derivations[6], can show that

$$\begin{cases} p_1 - c_1 = \frac{1}{\alpha s_0} \\ p_2 - w = \frac{1}{\alpha s_0} \end{cases} \tag{3.4}$$

i.e., the margins from two products are the same. In particular, it's one over α times market share of outside option, which is a function of (p_1, p_2).

[6] for details, see appendix A.1

Step 2: solve $\frac{\partial p_2}{\partial \delta_1}$ using implicit function theorem (IFT)

Using equation [4], we get the implicit function

$$F(p_2, w, \delta_1, \delta_2) = p_2 - w - \frac{1}{\alpha}(1 + exp(\delta_1 - \alpha p_1) + exp(\delta_2 - \alpha p_2)) \tag{3.5}$$

The implicit function theorem says that

$$\frac{\partial p_2}{\partial \delta_1} = -\frac{F_{\delta_1}}{F_{p_2}}$$

After some more derivations[7], can show that

$$F_{\delta_1} = -\frac{1}{\alpha}\frac{exp(\delta_1 - \alpha p_1)}{1 + exp(\delta_1 - \alpha p_1)}$$

and

$$F_{p_2} = 1 + \frac{exp(\delta_2 - \alpha p_2)}{1 + exp(\delta_1 - \alpha p_1)}$$

Thus

$$\frac{\partial p_2}{\partial \delta_1} = -\frac{F_{\delta_1}}{F_{p_2}} = \frac{1}{\alpha}\frac{exp(\delta_1 - \alpha p_1)}{1 + exp(\delta_1 - \alpha p_1) + exp(\delta_2 - \alpha p_2)} = \frac{1}{\alpha}s_1 \geq 0 \tag{3.6}$$

I.e., when fixing w, non-store brand price p_2 indeed goes up with store brand preference δ_1. Furthermore, the slope is proportional to store brand market share: the more popular is the store brand, the more p_2 goes up. It makes sense because when the store brand takes up a higher market share originally, the retailer can gain more from raising its price p_1. This gives the retailer more incentive to raise the non-store brand price p_2.

3.3.3 Bargaining Effect

Recall that bargaining effect refers retailer forcing down the wholesale price w due to enhanced bargaining position with a higher store brand preference δ_1. Therefore in this subsection, I characterize $\frac{\partial w}{\partial \delta_1}$, i.e., how does w change in response to δ_1. The idea is the

[7] for details, see appendix A.2 and A.3

same: I first derive the first order condition of w as a Nash bargaining solution. This gives an implicit function of (w, δ_1, δ_2). Because of the simultaneous move assumption, $\frac{\partial p_1}{\partial w}$ and $\frac{\partial p_2}{\partial w}$ will be zero. At last I use the implicit function theorem to derive $\frac{\partial w}{\partial \delta_1}$.

Step 1: w as Nash bargaining solution

Following Nash bargaining literature, w maximizes the Nash product

$$np(w, p_1, p_2) = (\Pi_r(w, p_1, p_2) - D_r(p_1))^{b_r}(\Pi_m(w, p_1, p_2) - D_m)^{1-b_r}$$

therefore the first order condition: $\frac{\partial np}{\partial w} = 0$. In the simple case I set $b_r = \frac{1}{2}$, then

$$\frac{\partial np}{\partial w} = \frac{\partial \Pi_r}{\partial w}\Pi_m + (\Pi_r - D_r)\frac{\partial \Pi_m}{\partial w} = 0$$

since

$$\frac{\partial \Pi_r}{\partial w} = \frac{\partial}{\partial w}[(p_1 - c_1)s_1 + (p_2 - w)s_2] = -s_2$$

and

$$\frac{\partial \Pi_m}{\partial w} = \frac{\partial}{\partial w}(w - c_2)s_2 = s_2$$

we have

$$\Pi_r - D_r = \Pi_m \tag{3.7}$$

i.e., the additional profit to retailer R from agreement is the same as the additional profit to manufacturer M. (recall the disagreement value to manufacturer, D_M, is assumed to be zero)

Step 2: solve $\frac{\partial w}{\partial \delta_1}$ using IFT

From equation [7] we get the implicit function

$$G(w, \delta_1, \delta_2) = \Pi_r(\delta_1, \delta_2) - D_r(\delta_1, \delta_2) - \Pi_m(w, \delta_1, \delta_2) \tag{3.8}$$

And IFT says that

$$\frac{\partial w}{\partial \delta_1} = -\frac{G_{\delta_1}}{G_w}$$

To simplify notations, let $N_1 = exp(\delta_1 - \alpha p_1)$, $N_2 = exp(\delta_2 - \alpha p_2)$, and $D = 1 + N_1 + N_2$, then $s_1 = \frac{N_1}{D}$, $s_2 = \frac{N_2}{D}$. Then after some more derivations[8], can show that

$$G_w = -s_2$$

and

$$G_{\delta_1} = -\frac{D^2}{\alpha(1+N_1)^2} s_2^2 s_1 (1+s_0)$$

Finally

$$\frac{\partial w}{\partial \delta_1} = -\frac{G_{\delta_1}}{G_w} = -\frac{D^2}{\alpha(1+N_1)^2} s_1 s_2 (1+s_0) < 0$$

So an increase in store brand preference δ_1 indeed lowers the wholesale price w. Here we also see that if s_1 is close to 1, then $s_2 = 1 - s_0 - s_1$ is close to 0, then $\frac{\partial w}{\partial \delta_1}$ is close to 0 as well; the wholesale price w is already very low before δ_1 increase, so it cannot be lowered by very much.

3.3.4 Total Effect

Now combine the two effects and derive the total effect of store brand preference δ_1 on non-store brand price p_2. Instead of fixing the wholesale price, w now is allowed to change, and it satisfies the first order condition [7]. Again I use implicit function theorem to derive $\frac{\partial p_2}{\partial \delta_1}$, only that the implicit function in this case is of $(p_2, \delta_1, \delta_2)$ rather than (w, δ_1, δ_2). I.e., having the same step 1 (deriving first order condition) as in section 3.3, the step 2 becomes

[8] see appendix A.4 and A.5 for details

Step 2: solve $\frac{\partial p_2}{\partial \delta_1}$ using IFT

From equation [7] we get the implicit function

$$H(p_2, \delta_1, \delta_2) = \Pi_r(p_2, \delta_1, \delta_2) - D_r(p_2, \delta_1) - \Pi_m(p_2, \delta_1, \delta_2) \tag{3.9}$$

and IFT says that

$$\frac{\partial p_2}{\partial \delta_1} = -\frac{H_{\delta_1}}{H_{p_2}}$$

After even more derivations[9], can show that

$$F_{\delta_1} \propto (1 + N_1)^2 - N_2$$

and $F_{p_2} < 0$.

Therefore, $\frac{\partial p_2}{\partial \delta_1} > 0$ if and only if $(1 + N_1)^2 > N_2$. Recall $s_1 = \frac{N_1}{D}$ and $s_2 = \frac{N_2}{D}$, this condition says if store brand market share s_1 is sufficiently large compared to non-store brand market share s_2, then the non-store brand price p_2 will raise in response to increase in store brand preference δ_1, i.e., the pricing effect dominates. This verifies the intuition outlined before: when s_1 is already close to 1, then first, the retailer cares more about raising the store brand price, and second, the wholesale price would be already very close to marginal cost and could hardly be even lowered.

3.3.5 Discussion

The model is extremely simple and a few extensions should be considered. In particular, what will happen when the retailer sells more than one non-store brands? What if there are more than one retailers? What if the manufacturer produces more than one product? And the last but not the least, what happen if the game is in Stackelberg fashion: the retail price is set after observing wholesale prices.

[9]see appendix A.5 through A.9 for details

More than one non-store brand

This is not too different from the case considered above. The difference now is that the retailer disagreement value with manufacturer M_1 is the profit from selling its own store brand and the product from the other manufacturer M_2. Therefore, the bargaining effect in this case should be smaller than in the baseline model, but the basic pattern (that $\frac{\partial p_2}{\partial \delta_1} < 0$ when s_1 is small, and > 0 when s_1 is big) should be preserved.

More than one retailer

Suppose there are two retailers R_1 and R_2. R_1, as in the baseline model, sells its own store brand product 1. Besides, both R_1 and R_2 sells the non-store brand product 2. Then facing the competition from another retailer R_2, now retailer R_1 cannot fully extract surplus from increase in the preference of its own store brand; if he raises (p_1, p_2) by too much, consumers will simply go to R_2 that sells the same non-store brand product 2. Therefore, with competing retailers, we might not even see positive $\frac{\partial p_2}{\partial \delta_1}$. This is actually the case in data.

Multi-product manufacturer, Stackelberg game

Modeling multi-product manufacturer and Stackelberg game are more complicated extensions that I want to spend some time working on. In the Stackelberg game, the wholesale prices w are negotiated first, both parties having perfect foresight of what the retail prices conditional on w. In this case, $\frac{\partial p}{\partial w}$'s are no longer zeros, complicating the computation.

As of multi-product manufacturer, let's assume there is one retailer R and one manufacturer M that has two product lines: product 2 and 3. If it's simply assumed that M negotiates its products separately and independently, and he maximizes the joint profit, then the bargaining effect is like that in the "more than one non-store brand" case: selling multiple products gives the retailer a better outside option even without its own store brand; the bargaining benefits from its store brand is lessened. Note that even though M owns both product 2 and 3, its position is weaker than a single-product firm because now M cannot coordinate the bargaining processes between the two products: it cannot threaten to withdraw both products if it is not satisfied with one of the wholesale prices. How about dropping the assumption that product 2 and 3 are bargained separately? As far as I know, there is not a theory on coordinated bargaining. So it will be great (yet difficult) to make some advances in this direction.

3.4 Estimation and Results

In this section I use the Homescan data to test how does store brand market share affect non-store brand prices. Recall the observation is at barcode (UPC) level. The baseline regression is

$$log(p_{jrt}) = \alpha + \beta_{pl} s^{pl}_{jrt} + \beta_{npl} s_{jrt} + \delta_{jr} + \epsilon_{jrdt} \tag{3.10}$$

where j stands for a product (defined by UPC code), r stands for a retailer, and observations are aggregated at monthly level. Here the prices and sales are aggregated at retailer level because of the evidence that most retailers set uniform retail prices across stores.[10] s^{pl}_{jrt} is store brand market share within retailer r at month t, for the product module that j belongs to; and s_{jrt} is the market share of product j. The parameter I'm interested in is β_{pl}. I also include product-retailer fixed effect (δ_{jr}).

To have a consistent estimate of β_{pl}, it's necessary to have a few (stringent) assumptions. First, s^{pl}_{jrt} and s_{jrt} need to be uncorrelated with ϵ_{jrdt}. But it is really unlikely since the error term contains the cost of product j, which is correlated with s_{jrt}. Second, market shares (own and private label) must enter linearly into the price equation. However, as we see section 3, marginal effect of private label market share should be nonlinear; β_{pl} more negative when s^{pl} is smaller. What's more, not only all the market shares must enter linearly, but all the rival non-store brand shares must have identical marginal effects. Otherwise it is necessary to include all the market shares (of products within the same module) on the right hand side.

[10]see DellaVigna and Gentzkow, "Uniform Pricing in US Retail Chains"

The estimates are presented in the first column of Table 1. The effect of s^{pl} on P is positive but small: one standard deviation (23%) increase in private label market share leads to only 0.87% increase in non-private label price.

Table 3.1: Linear Model Results

	(i) OLS, UPC level	(ii) OLS, module level	(iii) IV1	(iv) IV2
β_{pl}	3.8%***	10.1%***	-54.5%**	-32.8%
se	0.0016	0.013	0.23	0.65
IV1: median s^{pl}	X	X	√	X
IV2: lag s^{pl}	X	X	X	√
# obs.	7,540,957	348,872	70,992	68,034

3.4.1 Endogeneity

As discussed above, the market shares are likely to be endogenous. So in this subsection I try to find instruments for it. Here rather than including market shares of all products, I aggregate price and sales data of all non-store brand products. Thus, due to collinearity of s^{pl} and s^{npl} (npl stands for non-store brand), I have only s^{pl} on the right hand side:

$$logP_{jrt} = \alpha + \beta_{pl} s^{pl}_{jrt} + \delta_{jr} + \epsilon_{jrt} \tag{3.11}$$

where j stands for a product module. This aggregation is important because instead of finding IV for market shares of all products, now I only need IV for s^{pl}.

Compared with regression at barcode level, this regression is more hand-wavy about what's really going on. For example, say an increase in store brand market share leads to high-end product exit, then this regression will estimate a negative β_{pl} even if surviving non-store brands don't lower price; it doesn't distinguish product entry/exit with raising/lowering prices. But the barcode level regression [10] captures only the pricing effect. So I first compare the estimated β_{pl} from [11] to that from [10] (barcode level data).

The estimates are presented in column 2 of Table 1. First note that the number of observations drops if more than seven million to just over 300 thousand.[11] Second, the estimated β is larger than that with barcode level data. As discussed above, this can be due to product entry and exit. But the message from these two regressions are similar: a higher store brand market share has a positive yet small effect on non-store brand price.

IV1: retailer level store brand market share

The first IV for s^{pl} is a measure of store brand popularity at retailer level within the same period. The idea is simple: it's possible that when the retailer advertises, it advertises for all its private labels: Target, for example, advertises Archer Farms. This leads to an increase in the share of all Archer Farms brands independent of a price shock within a specific category. In addition, tastes for all private label products within a retailer are likely correlated: if a consumer tries Archer Farms maple syrup and really likes it, then this experience might also raise his preference for Archer Farms snacks.

In particular, I use the median of s^{pl}_{jrt} private label market share at retailer r at time t to instrument s^{pl} for a particular j.[12] The estimates are presented in column 3 of Table 1. Having an instrument completely changes the estimate; it reverse the sign of β_{pl}. Now a 23% increase in store brand market share is estimated to lead to a -13.0% price drop of non-store brand products. Linking to the model in section 3, this means the bargaining effect dominates the pricing effect. Also note that since now I drop all observations with zero store-brand share and only keep a balanced panel, the number of observations drops to only over 70 thousand.

[11] How many number of products a product module has? It turns out to be quite skewed: the mean is 23.4 but the median is only 6. A systematic way to deal with this skewness is needed.

[12] For a lot of product modules, $s^{pl} = 0$ throughout the time. Estimating β_{pl} on these modules seems meaningless. Besides, if $s^{pl} = 0$, then store brand advertising and correlated taste for store brands will not affect those modules. Therefore I drop all observations with $s^{pl} = 0$, and construct iv1 as median of s^{pl}_{jrt}, conditional on $s^{pl}_{jrt} > 0$. I further restrict to a balanced sample. This is because modules with full-period private label presence are more used to private labels, thus the setting could be better modeled with equilibrium. I.e., product entry/exit and learning tend to be less intense in those modules.

A note on the validity of IV1. The underlying model is that

$$logP = f(s^{pl}) + c + e \tag{3.12}$$

For IV1 to generate consistent estimates, I need (at least) two assumptions. First, the error term, $\epsilon = c + e$, has only cost shocks but not demand shocks. Second, market share enters linearly, i.e., f is a linear function. I will talk about nonlinear f in the next subsection, but for now, I'll use another IV to (try to) verify the validity of IV1.

IV2: lag store brand market share

The second IV for s^{pl}_{jrt} is the lag of s^{pl}_{jrt}. Note that s^{pl}_{jrt-1} may not be a valid IV if the cost, c, is an AR(1) process. In particular, assume that $c_t = \rho c_{t-1} + v_t$, then model [12] becomes

$$logP_t = \rho logP_{t-1} + f(s^{pl}_t) - \rho f(s^{pl}_{t-1}) + v_t + e_t - \rho e_{t-1} \tag{3.13}$$

According to this model, then a valid IV would be $(s^{pl}_{t-2}, s^{pl}_{t-3}, ...)$. But note that shares of store brand months ago could be noisy predictors of current market share, and as lag increases, this IV becomes more and more noisy.

Results in column 4 of Table 1 is obtained when I use $(s^{pl}_{t-2}, s^{pl}_{t-3})$ to instrument (s^{pl}_t, s^{pl}_{t-1}). It is insignificant, but it is the same sign as the estimate of β_{pl} using IV1, and the levels are similar. This is encouraging sign that IV1 could actually work. I could also include longer lags, e.g., s^{pl}_{t-4} and s^{pl}_{t-5} as well to instrument (s^{pl}_t, s^{pl}_{t-1}). This estimates β_{pl} to be -47.1%, but still insignificant.

3.4.2 Nonparametric Estimation

The model in section 3 predicts β_{pl} to be more negative when the share of store brand, s^{pl}, is close to zero. This subsection tries to capture this nonlinearity through nonparametric estimation. In particular, I split the sample into bins based on s^{pl}, and then estimate equation [11] separately on different bins.

To be specific, I first split the sample into two bins. The first bin has observations with low s^{pl} ($s^{pl} < 36.3\%$), and the second bin has observations with high s^{pl} ($s^{pl} \geq 36.3\%$). The results are presented in column 1 and 2 of Table 2. ($\beta^L_{pl}, \beta^H_{pl}$) corresponds to the first and the second bin, respectively. We see that for smaller s^{pl}, β_{pl} is estimated to be -135%, while for larger s^{pl}, β_{pl} is estimated to be -93%. This is consistent with the theory prediction that store brand market share has a more negative effect on prices when it is smaller.

I also tried to split the sample into more than 2 bins. The results with 3 bins are presented in column 3, 4, and 5 of Table 2. The bin with the lowest s^{pl} still gives the most negative β_{pl}, and even more negative than that with 2 bins. This is expected since with 3 bins, s^{pl} are even lower than that with 2 bins. Besides, bin 2 and bin 3 now give less significant estimates (only significant at 10% level). This is due to the smaller sample size. If I split the sample into more than 3 bins, then only the first bin generates significant result. I also present the result from bin 1 (lowest s^{pl}) with 4, 5, 6 bins in column 6, 7, 8 of Table 2, respectively. As expected, the estimate becomes more and more negative as s^{pl} becomes smaller.

Table 3.2: Nonlinear Model Results

# bins	2		3			4	5	6
s^{pl}	.01-.36	.36-1	.01-.25	.25-.49	.49-1	.01-.20	.01-.17	.01-.15
β_{pl}	-1.35**	-0.93***	-2.85**	-1.19*	-1.62*	-4.02**	-5.52**	-8.46*
se	0.65	0.52	1.3	0.69	0.93	1.7	2.3	4.3
# obs.	35,496	35,496	23,664	23,664	23,664	17,748	14,199	11,832

3.5 Conclusion and Future Steps

In conclusion, I use a simple model to show that there are two channels through which store brand market share can affect non-store brand prices. The pricing effect refers to that once consumers like the store brand more, the retailer wants to take advantage of it by raising store brand price. Since the retailer has some market power (e.g., due to travel cost), he internalize some of the business stealing externalities and raise the non-store brand price as well. The bargaining effect refers to that a stronger store brand gives the retailer a better disagreement value with non-store brand manufacturer: without the non-store brand, the retailer could still retain a large share of consumers. This drives down the wholesale price.

The combined effect is nonlinear: when the market share of store brand is originally small, the pricing effect is minimal and the bargaining effect tends to dominate. When the market share of store brand is already high, then the pricing effect is strong and the bargaining effect is minimal: wholesale price is barely above manufacturing cost already. Therefore, store brand market share has a more negative marginal effect on non-store brand prices when it is smaller.

I then tested this nonlinear pattern with Nelson Homescan data. To deal with endogeneity of market share, I instrument it with median store brand share at retailer level. The logic is that when the retailer advertises its store brand, preference of all store brand products raise. In addition, it's conceivable that preference for the same store brand is correlated, while cost shocks are not. I further test the validity of this IV with another IV, lags of the store brand market shares. Both these IVs generate similar results, that a higher store brand market share drives down significant the non-store brand prices. I also found some evidence of the nonlinear effect through splitting the data into bins based on store brand market share: in the subsample with smaller store brand market share, the marginal effect is more negative.

Future steps

Separating bargaining effect in data

The first thing I want to do is to actually separate bargaining effect from pricing effect in the data. One thing I could do is to first estimate a demand system, and then use first order conditions to back out wholesale prices. Then I can directly check how does market share of store brands affect wholesale prices. Though this is the approach taken by a few empirical papers (e.g., Ellickson et al (2017) [16], Draganska et al (2010) [14]), it'll be better if it can be done through finding some patterns in the data. One thing that comes to mind is that the magnitude of bargaining effect depends on supply side market structure. If the supply side is dominated by a few firms, then they have power to raise the wholesale price above manufacturing cost. However, if the supply side if nearly competitive, then the wholesale price will be close to manufacturing cost, leaving little room for the bargaining effect.

Look at product entry/exit

I also want to include product entry and exit into the model. It is interesting for a few reasons. First, store brand entry is yet another reason for higher private label market share, providing another potential IV for s^{pl}. However, entry and exit decisions are hardly exogenous. This brings about an even more interesting (and hard) question of modeling product entry and exit decisions. For example, does a higher store market share lead to more or less non-store brand entry and exit? What are those products that enter and exit the market? What are the welfare consequences?

Chapter 4

The Value of Market Intelligence Data in Online Marketplace: a Case Study of Taobao Lipstick Sellers

4.1 Introduction

There is a fundamental change in the role that data and information play in todays economy due to the technological progress in data collection, storage, and analysis. Firms large and small invest heavily in data technology and data science in order to make more informed decisions.[1] However, there has been little work about how data and information affect firm decision-making, and what their impact is on market outcomes.

To fill this gap, we look at one important kind of data for business strategies: the market intelligence data. It stands for the information relevant to a company's market: trends and competitor and customer monitoring.[2] We study an online sales platform, Taobao, that provides its sellers with a market intelligence tool called Market Insight. Market Insight allows its subscribers to see what other sellers are selling, and how much money they make from each product.[3] By telling subscribers which products are profitable and which are not, Market Insight could influence these sellers' product assortment decisions, which in turn determine market outcomes such as competition, product variety, total sales on the

[1] A recent survey found that businesses globally are spending close to $40 billion annually on technology and services for data analytics, increasing by 12% each year. "Is Your Business Masquerading as Data-Driven?", Harvard Business Review, 2020.

[2] Cornish, S. L. Product Innovation and the Spatial Dynamics of Market Intelligence: Does Proximity to Markets Matter? Economic Geography. Volume 73, Issue 2 (April 1997).

[3] Taobao obtained permission to use this information from participating sellers.

platform, and consumer welfare.

We ask three interrelated questions. First, we try to understand how Market Insight affects online seller decisions, and its impact on seller performance. Second, we turn to the impact of Market Insight on market outcomes. In particular, we assess the value of Market Insight for the e-commerce platform and for consumers. Finally, we solve the optimal design problem that maximizes the platform sales,[4] and then compare it to the socially optimal design that maximizes consumer welfare.

I focus on a sector where Market Insight is likely valuable: lipstick. On the one hand, figuring out the expected revenue of a particular lipstick product is difficult—the lipstick industry is characterized by large changes in fashion and styles. On the other hand, deciding which lipstick brands and models to carry is an important consideration for online sellers because they typically carry only a few out of hundreds, if not thousands, of different lipsticks.[5]

To identify the value of Market Insight to online lipstick sellers, we exploit a change of Market Insight design in August 2019, after which some subscribers received more precise information through observing twice as many other sellers. We find that this increase in information led to 40% higher revenue for treated sellers. To further investigate the mechanism, we first document that the policy change significantly affects seller product choice: the treated sellers' rate of introducing a new lipstick rose by 20%. To see what these newly introduced lipsticks are, we estimate the time-varying profitability[6] of each lipstick brand and document that this change in information induces treated sellers to sell more profitable lipsticks.

[4]In this study we assume the e-commerce platform's objective is to maximize total sales on the platform. This is because the main income of the platform comes from advertisement spendings of online sellers, which is proportionate to the total sales on the platform (from readings of other platforms).

[5]Lipstick sellers in this study are small retailers who buy from wholesalers and sell to end consumers. Unlike lipstick manufacturers like Dior or Chanel, these sellers can choose which lipstick brands and models to carry at any given time.

[6]As will be formally defined in the model section, profitability is a single index summarizing the expected profit of the brand, excluding seller abilities (e.g., seller reputation that affects demand).

To assess the value of Market Insight to consumers and to the online platform (Taobao), and to determine how to optimally design Market Insight, we build a structural model that links the information contained in Market Insight to seller product choice and ultimately to market outcomes such as product variety, total platform sales, and consumer welfare. Specifically, sellers make product-assortment decisions combining their prior about how their lipstick would sell and the information from Market Insight; the more sellers they observe through Market Insight, the more precise information they get.

We assess the value of Market Insight to consumers and to the platform by simulating what products will be offered, who will offer them, and at what price, if Market Insight did not exist. In this case, sales drop by 8%, and consumer welfare drops by 0.7%.

To solve the platform-optimal information design of Market Insight, we simulate product varieties on the platform if Market Insight allowed a seller to follow 0 to 400 other sellers. Sales and consumer welfare first increase and then decrease with the amount of information provided in Market Insight. This nonmonotonicity is caused by the fact that seller objectives differ from those of the platform. When sellers become more aware of which products are less profitable, they drop those products, thus reducing product variety on the platform. This lowers consumer welfare since a consumer might not be able to find the product she wants. As a result, this reduces the total number of transactions on the platform, and ultimately lowers platform sales.

For the lipstick sector, sales would be maximized by providing three times more sellers to follow. Sales would increase another 8%, and consumer welfare would rise 0.8%. Finally, we compare the platform-optimal policy with the social-optimal policy (which maximizes consumer welfare) to find that they result in very similar outcomes for both consumers and the platform, showing that in this case, the platform acts like a benevolent social planner.

Related Literature

Social learning has drawn theorists' attention for a long time. For an overview, see Golub and Sadler (2017). Some empirical papers document evidence for social learning. Griliches

(1957) , Foster and Rosenzweig (1995), and Conley and Udry (2010) considered the adoption of new technology and social learning in agricultural settings. In industrial organization literature, a few recent papers document evidence for social learning about production technology in the oil industry (Kellogg (2011) , Covert (2015) , Hodgson (2018) , and Steck (2018)). In contrast, this paper documents evidence for social learning in a very different context, that of product assortment decisions in an online retail market.

This study also contributes to the burgeoning information design literature. Information design refers to a setting where party A, an "information designer," controls the information structure (rather than the information itself), while party B chooses optimal actions given the information structure, which affects party A's payoff.[7] For an overview, see Bergemann and Morris (2017). Numerous theoretical papers on this subject have applications in areas such as grade disclosure and matching markets (Ostrovsky and Schwarz (2010)), voter mobilization (Alonso and Camara (2016)), and traffic routing (Das, Kamenica, and Mirka (2017)). While the literature provides plenty of insight from theory, empirical tests are rare. One example is Asquith et al. (2019), which studies the effect of mandatory dissemination of price and volume information for corporate bond trades.

This paper also contributes to the thriving literature on two-sided markets—a heated topic in the past decade. For reviews, see Rysman (2009) and Rochet and Tirole (2006), which address the body of literature whose focus has been on platform pricing strategy; the literature largely ignores information strategy. Platform information strategy is important because platforms collect huge amounts of data, and their information policies significantly affect all parties involved. Information strategy is also interesting because, as with pricing strategies, outcomes are determined by the interaction between the two sides of the platform. As far as we are aware, this study marks the first attempt to study platform information design.

Last but not least, this paper contributes to the small volume of literature on how big

[7] In my setting, the platform chooses, for example, how many rival sellers an informed seller may follow (information structure) to affect behaviors of informed and uninformed sellers on the platform, which affects outcomes of interest to the platform, such as total platform sales, competition, and consumer welfare.

data affects firm performance. A few empirical works on the "scale effect" of data address how increases in the amount of data can aid prediction. Varian (2014) and Lambrecht and Tucker (2017) suggest that the scale effect may not be very helpful in improving prediction performance, while De Fortuny et al. (2013) find that the performance of a variant of a naive Bayesian classification algorithm continues to improve even after the number of data entries exceeds some very large amount (500 million in some cases). Bajari et al. (2019) finds that while the number of products has a flat effect, the number of time periods improves forecast accuracy. My setting addresses the other aspect of big data, that data is collected in real time. This high-speed feature is important in this setting since the lipstick sellers need to know the market trends in time; knowing, say, consumers wanted Dior 999 a year ago isn't helpful. As far as we are aware, this is the first empirical study to address the impact of real-time data on firm performance.

The paper is organized as follows. Section 2 describes the background and the data. Section 3 presents detailed evidence of how Market Insight affects sellers. Section 4 presents the model. Section 5 describes estimation and identification. Section 6 calculates the value of Market Insight to consumers and the platform, and then solves the platform-optimal design problem. We conclude in Section 7 with a discussion of future research.

4.2 Background and Data

This section describes Taobao and the market intelligence tool, Market Insight, and the policy change the occurred within Market Insight in August 2019. We also describe the lipstick industry, discussing why we focus on it, who the sellers are, and what they do. Lastly, we describe our data.

4.2.1 Taobao and Market Insight

Taobao is by far the largest e-commerce platform in China; it has a market capitalization of more than 800b USD, about half that of Amazon. Its main competitors in China, JD and PDD, both have market capitalizations of less than one-sixth that of Taobao.

Taobao offers a market intelligence tool named Market Insight to its sellers. A subscriber of Market Insight will be able to i) observe the best-selling products in their category and ii) see what other sellers in the category are offering, what prices are being charged, and how many units are being sold.

The Policy Change of Market Insight

Market Insight has two versions. The "basic" subscription allows a subscriber to observe up to seven other sellers. The "pro" subscription had permitted observing up to 30 sellers until 2 August 2019, when a policy change raised the number of observed sellers to a maximum of 60. This exogenous policy change, while not affecting basic subscribers, offered pro subscribers substantially more information.

4.2.2 Lipstick Industry and Taobao Lipstick Sellers

Demand for lipstick is substantial and steadily growing in China. Lipstick is the largest category within cosmetics on Taobao; the average female user of Taobao buys four lipsticks per year. From 2016 to 2019, the lipstick sector experienced a manyfold increase in total revenue.

More importantly, Market Insight data is expected to be exceptionally valuable in this product category, because lipsticks see large and frequent demand shocks due to rapidly shifting fashion trends. Big brands bring out new colors, new product lines, and limited editions every few weeks.[8] Because of the rapid shifting fashion trends and wide product diversity in terms of brands and colors, consumers might not know every product very well. Influencers specialize in promoting new colors and new product lines, generating shocks

[8] New brands appearing and entry from brands in other industries are frequent. In my sample period, many Asian (e.g., Chinese, Korean, Thai) brands emerged on Taobao, many of which managed to become top sellers for at least some months. Targeting younger generations and rural areas, they are typically cheaper than well-known brands like Dior.

in consumer demand.[9] Celebrities and television shows also produce demand shocks: one day a main character in a popular show wears a particular lipstick color, and the next day viewers are wearing it.

Another reason Market Insight data might be especially valuable to this category is that Taobao lipstick sellers are very small retailers, offering a limited number of products. They buy from wholesalers and sell directly to end consumers.[10] For example, the so-called "Daigou" purchase lipsticks from cosmetic stores overseas and then ship to buyers in China.[11] According to the data, the median Taobao lipstick seller carries only two items,[12] likely because these small retailers cannot afford the storage costs or the formidable working capital required to offer a variety of products across multiple brands. These sellers, therefore, are faced with the challenge of selecting a few lipsticks out of hundreds or thousands, without knowing which are likely to be most profitable.

4.2.3 Market Insight Subscribers

Overall, about 1% of all sellers subscribe to Market Insight Pro, and 3% subscribe to Market Insight Basic. Subscribers to Market Insight pro are the largest sellers in every measure: they have the highest average monthly revenue, the highest rate of having Gold and Global status, the largest fan base, and the most stars. Basic subscribers are smaller than pro sellers, but larger than nonsubscribers.

[9] Jiaqi Li, the "Lipstick Prince of China," has more than 37 million followers on his TikTok account; everything he recommends sees immediate jumps in sales, often leading products to be sold out within a few days.

[10] Alibaba group has two large platforms, Taobao and Tmall. While Tmall is the B2C platform with manufacturer sellers such as Dior and Chanel, Taobao is the C2C platform.

[11] The Daigou have arisen as more Chinese nationals study and live overseas. Because foreign companies typically charge higher price in China than abroad (Dior, for example, engages in third-degree price discrimination), overseas pricing becomes favorable. In addition, many lesser-known brands don't have an official online store; consumers will have to buy it from overseas.

[12] An item is typically several (up to 20) lipsticks of different colors within the same brand. This is how Taobao sellers report what products they carry. Admittedly, while there's a fat tail of small sellers, even the top fifth percentile of sellers (in terms of their revenue) carry only five items on average.

4.3 Impact of Market Insight on lipstick sellers

This section provides reduced-form evidence that Market Insight is valuable to lipstick sellers. Exploiting the policy change, we find that Market Insight has a positive impact on seller performance and significantly affects seller product choice.

4.3.1 Diff-in-Diff

We ran the following diff-in-diff regression

$$y_{kt} = \gamma_1 TREAT_{kt} + \gamma_2 PLACE_{kt} + \xi_k + \tau_t + e_{kt} \tag{4.1}$$

where k is seller and t is month. y_{kt} measures either seller performance or product choice. The parameter of interest is γ_1. $Treat_{kt} = 1$ if and only if seller k is a Market Insight pro subscriber and month t is after August 2019; otherwise $Treat_{kt} = 0$. Therefore, γ_1 is the causal effect of the policy change on Market Insight pro sellers. There is also a seller fixed effect ξ_k to absorb any level difference across sellers and a month fixed effect τ_t to account for seasonality.

To check for robustness, $Place_{kt}$ indicates seller k is a Market Insight basic subscriber and month t is after August 2019. Thus γ_2 is the placebo effect of the policy change on basic subscribers, and it functions to rule out any common trend shared by pro and basic sellers. For example, suppose Dior comes into higher demand from August through October, coinciding with the timing of the policy change. In this case, since both basic and pro sellers have more information, they might be more likely to notice this trend.[13] If the policy change has no impact on basic subscribers, this rules out common time trends shared by basic and pro sellers, such as those coming from demand shocks.[14]

[13] Pro sellers should experience a larger effect because they have more information than basic sellers, but it is still the case that this event should affect both.

[14] For this regression, we keep a balanced sample of sellers to eliminate noise from seller entry and exit. We also exclude switchers of Market Insight subscription: any seller remaining in the sample is either a pro seller, a basic seller, or a nonsubscriber throughout the sample period. Therefore,

4.3.2 Seller Monthly Revenue

Is Market Insight valuable to a lipstick seller? To answer this question, we look at the impact of the policy change on log(seller monthly revenue + 1).[15] The result is shown in Table 1, column 1. $\gamma_1 = 0.4$ means that on average, the policy change raises treated (Market Insight pro) sellers' monthly revenue by 40%.[16] Also notice that γ_2 is essentially 0, ruling out any common time trend shared by pro and basic sellers.

Table 4.1: Diff-in-Diff Effect on Seller Performance and Product Assortment

	Mon. revenue	Rate of intro.
TREAT	0.39***	0.06***
	(0.08)	(0.02)
PLACE	−0.04	0.01
	(0.05)	(0.01)
Switcher	N	N

- Column 1 is effect of policy change on seller monthly revenue
- Column 2 is the effect on rate of introducing new products
- Robustness error clustered at seller level (observation at seller-month level)

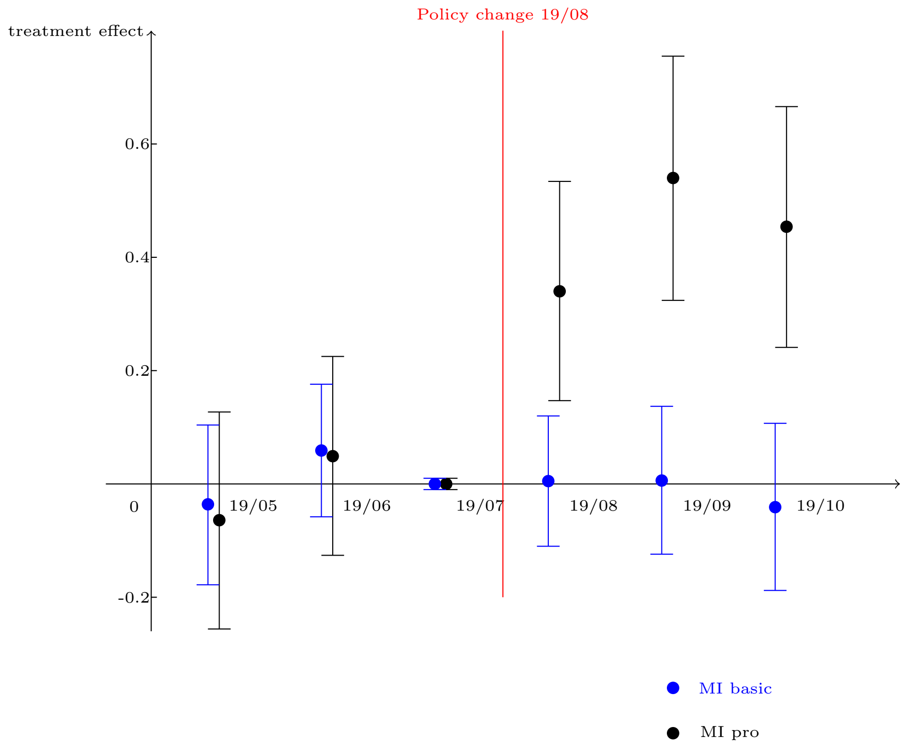
I also plot the monthly treatment effects on treated (γ_1) and placebo (γ_2) sellers in Figure 1. Before the policy change, there is no pretrend for pro sellers: they share similar time trends as nonsubscribers. However, immediately after the policy change, treated sellers see a significant jump in monthly revenue.

all the variation in $Treat_{kt}$ should come from the policy change. To have more treated sellers, we only use a subsample three months before and three months after the policy change; that is, from May 2019 through October 2019. Using the whole sample doesn't change the results; see appendix B.2. Entering and exiting sellers account for 40% of all sellers who ever sell lipsticks during the six months. Among the remaining sellers, nonswitchers account for 97%.

[15]Monthly revenue is well-measured and widely used in Taobao (as in many other settings, seller cost is not observable, so there is no way to directly measure seller profit). The "+1" is there to account for zero-sales, which is common for online sellers since they are frequently quite small.

[16]This effect may seem strikingly large, but in speaking with people at Taobao, they confirm this is within a reasonable range; since the Taobao sellers are quite small, a 40% increase might represent only a few hundred dollars.

Figure 4.1: Event Study, Seller Monthly Revenue



Note: monthly diff-in-diff treatment effects (compare to non-Market Insight sellers), normalized at 2019/07. The dots represent the point estimates, and the bars represent the 95% confidence interval.

The monthly treatment effects on placebo (basic) sellers are close to zero and insignificant both before and after the policy change. Again, this rules out any common time trend shared by basic and pro sellers. In addition, it is also comforting to know that the basic sellers have almost exactly the same time trend as nonsubscribers both before and after the policy change.

4.3.3 New Product Introduction

In this section we present evidence that information affects seller product choice. Intuitively, Market Insight should help sellers make better assortment decisions because it reveals which products are making money and which are not for other sellers.[17] Among various methods to measure product choice, here we use the rate of introducing a new lipstick for each seller. We focus on this method because it is a robust measure of seller assortment, which is particularly helpful because there's no measure supplied by Taobao about how profitable each product is (which we will construct later). This measure is also widely used by Taobao.

Continuing to exploit the policy change, now y_{kt} in equation [1] is an indicator whether seller k introduces any new lipstick in month t. As shown in Table 1, column 2, the policy change raises the estimated introduction rate by nearly six percentage points. Because the average introduction rate among pro subscribers is about 30%, the policy change raises this by nearly 20%.

Again, the placebo group of basic sellers doesn't react to the policy change, ruling out a common time trend shared by all Market Insight sellers. This suggests the result is not driven by some big new lipstick debut that happened to coincide with the policy change.

[17] There are other channels through which information can help, such as item title improvement. Through Market Insight a seller can see keywords that help attract traffic. For example, the keyword "Rouge Dior 999, Jiaqi Li recommended" might be very popular, and a seller of Rouge Dior 999 might want to include the keyword "Jiaqi Li recommended" in the title. This wording information, however, does not affect product choice; the only information in Market Insight that should influence product assortment is product sales.

Figure 4.2: Event Study, Rate of Intro.

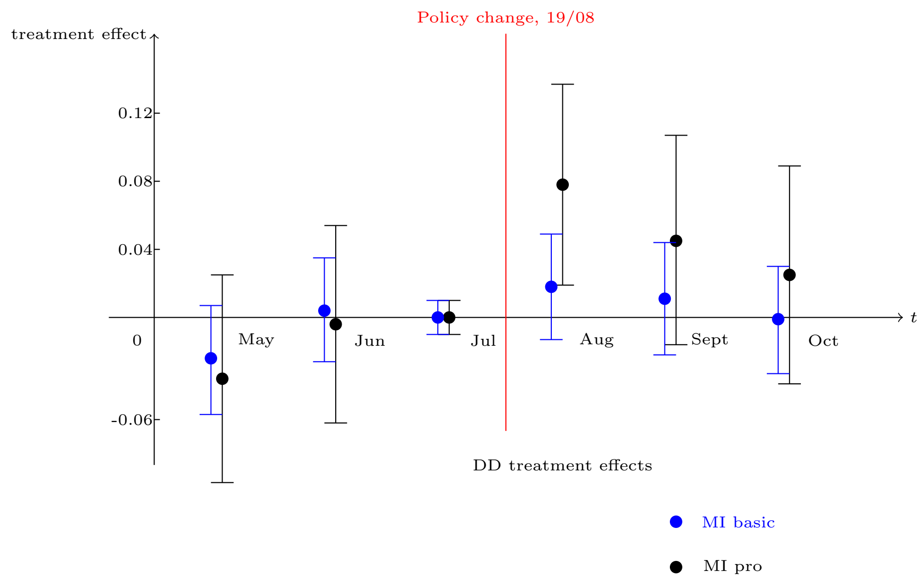


Figure 2 plots the monthly treatment effects for treated (pro) and placebo (basic) sellers. The monthly treatment effects for treated sellers are significant only in August, the month immediately following the policy change; in September and October, the treatment effects, though not significantly different from zero, are still large in magnitude. One cause for the insignificant estimates is the weaker power in this regression: a seller only introduces a new lipstick every few months, so if a new item is introduced in month t, there is unlikely to be another new item introduced in month t+1; identifying the treatment effects for each month might therefore be asking too much from the data. The diff-in-diff specification in equation [1] estimates the aggregate treatment effect for months August through October; with enough power, the treatment effect is significant at the 1% level. Basic sellers share the same time trend with nonsubscribers both before and after the policy change; the estimated monthly treatment effects are never significant in any single month, nor jointly significant after the policy change.

4.3.4 Expected Revenue of the Newly Introduced Products

What are those newly introduced products? Since subscribers of Market Insight see revenue made from different products by other sellers, these new products should have higher revenue. To check the validity of this story, we need a measure of the expected revenue of each product at each time period. There are two challenges in constructing such measure: first, while a product should be a combination of brand, color, and texture, Taobao has trouble reliably collecting color and texture information. Second, the revenue of a specific seller-product pair depends on both how attractive the seller is and how attractive the product is; it is nontrivial to separate the seller component and the product component. In section 4.2 we will make assumptions on the seller and product components. Then we construct a measure of the expected revenue of each product by excluding the seller component.

We find Market Insight indeed helps a seller to find better products to carry in terms of expected revenue. If a lipstick's expected revenue goes up by one standard deviation, after the policy change treated sellers rate of carrying that lipstick goes up by 13% (see section 4.2 for the definition of a product's expected revenue, and appendix B.1 for regression details).

4.3.5 Robustness

The identification assumption in the diff-in-diff analysis is that all sellers, regardless of Market Insight usage, share the same time trend. For example, when platformwide sales surge in November, the assumption says that the sales boom is proportionately distributed among all sellers. One problem with this is that larger sellers could gain disproportionately more of the sales boom[18]—and pro sellers tend to be larger. Basic sellers are also larger than nonsubscribers, but they are considerably smaller than pro sellers. If the time trend satisfies 1) the largest sellers enjoy the sales boom while smaller sellers don't, and 2) the timing of the sales boom coincides with the policy change, then the treatment effects identified above

[18]For example, higher demand in November might only benefit big sellers because they are more visible to consumers.

are no longer causal.

Synthetic control can alleviate this concern. Among nonsubscriber and basic sellers, some are comparable in size to pro sellers. With these placed in a control group and assigned a higher weight, the monthly revenue (rate of introducing a new lipstick) before the policy change is exactly the same in the "synthetic" control group as in the treatment group. The synthetic control approach yields very similar estimates: treated seller revenue goes up by 40% right after the policy change, and their rate of introducing new lipsticks goes up by about five percentage points. In addition, both of these treatment effects are statistically significant (see appendix B.2).

4.4 Model

Reduced-form analysis shows that Market Insight affects seller product choice. In this section, we develop a structural model that links the information contained in Market Insight and seller product choice,[19] and then aggregate product offerings of all sellers to get platform competition, total sales, and consumer welfare.

We develop this model for three reasons. First, we want to investigate further into the mechanism of how Market Insight affects product choice. In particular, we want to formally define the product profitability used in Section 3.4. Second, we want to calculate consumer welfare and platform welfare given the product offering on the platform. Third, to calculate the value of Market Insight to consumers and to the platform, and to solve the platform-optimal (social-optimal) design of Market Insight, we need to simulate consumer welfare and platform welfare if the design of Market Insight were to change. The model makes all the three inquiries possible.

In particular, the model has three stages:

[19] In theory, Market Insight could also help the seller to make better pricing decisions. For example, if a seller sells Dior 999 and wants to learn what price generates the highest revenue, Market Insight could help with that. However, this is not the focus of this study. On the one hand, experiences from others are rather noisy signals. On the other hand, a Taobao seller could easily experiment with different prices: she could change price as often as she wants at no cost.

1. Sellers subscribe to Market Insight, basic or pro.
2. Sellers learn about profitability of each lipstick and then decide which lipstick(s) to carry; Market Insight affects this learning process.
3. Sellers set prices and compete in the product market.

The following sections describe each stage in more detail. First, we lay out a logit demand system and write down the marginal cost of each seller-lipstick combination and the optimal pricing rule. We also formally define the time-varying "profitability" of each product. Then, we use a Bayesian model to capture seller learning of product profitability, and importantly how Market Insight affects this learning process. Then, we describe how sellers make assortment decisions based on their information set. Finally, we describe how sellers select a Market Insight subscription.[20]

4.4.1 Demand

Consumer i's utility is given by

$$U_{ijkt} = \delta_{jkt} + \epsilon_{ijkt} \tag{4.2}$$

where j denotes a product (from this point forward a product is a brand),[21] k denotes a seller, and t denotes a month. ϵ_{ijkt} is distributed EV(0,1). δ_{jkt}, the mean utility of product j by seller k at month t, is given by

$$\delta_{jkt} = d_{jt} + \beta_d X_{kt} - \alpha log(p_{jkt}) + u_{jkt} \tag{4.3}$$

where d_{jt} captures time-varying attractiveness of product j at time t. $\beta_d X_{kt}$ captures how observed seller characteristics (Gold status, Global status, stars, and log(fans)) affect seller

[20] As explained later, we found significant noise within the data regarding seller Market Insight choice, even without modeling assumptions. This step, therefore, only exists to give a complete model of seller choice; it does not play a big role in later analyses.

[21] This is primarily due to data limitations. In the lipstick sector, while lipstick brands are well-documented by Taobao, the colors and textures are not. Therefore, in my estimation section I'm only able to use brand-level data; everything product-specific should be interpreted as brand-average.

demand. For example, according to Hui et al. (2018), we expect a positive and significant impact from Gold status, the most important certificate for Taobao sellers. In addition, u_{jkt} is an iid product-seller specific component of utility. One example is search ranking: given lipstick j, if seller k has a higher search ranking, then a j-k pair is more easily spotted by consumers, thus delivering a positive "convenience" utility u_{jkt}.[22]

4.4.2 Marginal Cost and Profitability

I assume the marginal cost m_{jkt} is given by

$$mc_{jkt} = \gamma_{jt} + c_k + \beta_c X_{kt} + v_{jkt} \tag{4.4}$$

where γ_{jt} is the cost state of product j at time t. It includes both the wholesale price and the marginal promotion cost. The wholesale price is the price paid by Taobao sellers to suppliers; for example, Tom Ford lipsticks might be more expensive than Dior, and thus have a higher γ_{jt}. The marginal promotion cost is the cost paid by Taobao sellers to other parties, including the platform. One example is the brand average pay-per-click (PPC) price—sellers pay this price to gain higher search rankings. Because the PPC price depends on who the sellers are and how fiercely they compete, it varies across brands and over time. v_{jkt} represents the seller-product specific effort. For example, if a seller bids a higher PPC price than rivals, then the v_{jkt} is positive. In this case, the seller also enjoys a higher search ranking. Therefore, v_{jkt} is correlated with u_{jkt}, causing an endogeneity problem of price. I also allow the observed seller characteristics X_{kt} to affect marginal

[22]Logit demand imposes strong assumptions on consumer substitution patterns, making any seller-product pair a distinct variety; substitution within a brand isn't necessarily higher than across brands. At first sight, this might not make sense: a buyer looking for Dior lipsticks might only choose among Dior sellers. But in the Taobao lipstick market, substitution isn't necessarily within same-brand sellers for three reasons. First, different sellers of Dior in many cases carry different Dior lipsticks: recall that lipsticks of same brand have a variety of colors and textures. A seller, on the other hand, typically only carries a few colors and textures. Second, a consumer could be looking for a specific color rather than a brand, therefore the substitution is actually within color. Third, for a consumer, different sellers are different. This is because the consumer-seller specific trust relationship is important in online markets; a consumer might only consider lipsticks offered by a trusted seller. In this case, the substitution is actually within seller.

cost.[23]

A key aspect is that c_k captures unobserved seller cost heterogeneity. While Taobao knows a lot about seller heterogeneity in attracting consumers and making sales (much of this heterogeneity is created and managed by Taobao, e.g., Gold status), this is not the case for seller cost—Taobao doesn't know where each seller obtains products nor at what price.

To solve for the seller's optimal pricing rule, we write down the predicted market share

$$log(s_{jkt}) = log(s_{0t}) + d_{jt} + \beta_d X_{kt} - \alpha log(p_{jkt}) + u_{jkt} \tag{4.5}$$

where s_{0t} is the share of the outside option. Solving for the FOC, the optimal pricing rule is

$$p_{jkt} = \frac{\alpha}{\alpha - 1} \times mc_{jkt} \tag{4.6}$$

Here, price is the marginal cost multiplied by a constant markup. Therefore, the profit is given by

$$\begin{aligned} log(\pi_{jkt}) =& log(M_{0t}) + log(s_{0t}) + (1-\alpha)log(\frac{\alpha}{\alpha - 1}) - log(\alpha) + \underbrace{d_{jt} + (1-\alpha)\gamma_{jt}}_{r_{jt}} \\ &+ (\beta_d + (1-\alpha)\beta_c)X_{kt} + (1-\alpha)c_k + u_{jkt} + (1-\alpha)v_{jkt} \end{aligned} \tag{4.7}$$

"Profitability" $r_{jt} \equiv d_{jt} + (1-\alpha)\gamma_{jt}$ is the time-varying product-specific index. Combining both demand and cost (d_{jt}, γ_{jt}), the profitability index represents how choosing product j affects profit, excluding seller components such as reputation and cost heterogeneity. For example, for any given seller, $r_{jt} = r_{j't} + 0.2$ means the expected profit of the seller carrying product j is 20% higher than that of carrying product j'.

[23] Taobao sellers could use their Gold status as a bargaining tool to lower their wholesale price, but Gold status also places more stringent requirements on product authenticity, driving up marginal cost.

4.4.3 Bayesian Learning about Profitability r_{jt}

Seller product assortment depends on what they believe about the profitability r_{jt} of each lipstick. In this section we describe how sellers form their beliefs regarding r_{jt} and how Market Insight influences this learning process.

I assume sellers learn about profitability r_{jt} from Market Insight for three reasons. First, profitability r_{jt} is important for a profit-maximizing seller since profit from a product is monotonically increasing in r_{jt}. Second, it is difficult for a seller to know every r_{jt}, which is a complex object containing ever-changing consumer demand, and wholesale and promotion costs that depend on seller competition. It is difficult for a seller to track all these components of hundreds of different lipsticks as they change over time. Third, Market Insight provides precisely this needed information. When a seller is interested in how much revenue is obtainable from product j, Market Insight allows the seller to observe how much revenue other sellers are getting from product j. This provides the seller with information about the expected revenue if she were to offer product j. Furthermore, observing more sellers can increase accuracy in estimating r_{jt}.[24]

I assume the sellers are Bayesian learners; they have a prior belief that $r_{jt} \sim \mathcal{N}(\mu_r, \tilde{\sigma}_r^2)$ for every product j, where μ_r is the mean of r_{jt}, and $\tilde{\sigma}_r^2$ is the variance of seller prior (or 1 over prior precision). Then each seller receives a free signal $\theta_{jkt,0}$ about r_{jt}. If she subscribes to Market Insight, the seller also gets a Market Insight signal $\theta_{jkt,MI}$ about r_{jt}. The seller forms the posterior belief about r_{jt} based on the prior, the free signal, and the Market Insight signal (if subscribed).

Note that a seller is assumed to have the same priors for different products. In particular, carrying the product in the past doesn't affect the current prior. There are three reasons for this assumption. First, recall that r_{jt} refers to brand-average profitability, while Taobao sellers typically carry only a very narrow selection of colors and textures within each brand.

[24]Other information such as competition is irrelevant conditional on product profitability. For competition, a seller thinking about selling Dior might want to know how many other sellers are selling Dior. However, information like the number of Dior sellers is not contained in Market Insight. In addition, since a seller only needs the competition information to determine expected profit, learning about competition generates no additional value conditional on profitability r_{jt}.

Therefore, carrying Dior 999 last month, for example, doesn't reveal the average profitability of Dior lipsticks this month. Second, profitability also changes rapidly over time due to large and frequent demand shocks. Finally, if carrying lipstick j in the past has information value for the seller, we might expect frequent experimentation with different products. But this is not the case in the data: sellers typically introduce less than one new brand in a three-month period.

Also note that we don't restrict the variance in the prior of r_{jt}, $\tilde{\sigma}_r^2$ to be the same as σ_r^2 - the true variance in r_{jt}; this is because sellers could be ill-informed. In addition, different sellers share the same prior. Here we abstract away from differences in seller sophistication and experience, although in reality, one would expect a bigger seller with more experience to have a $\tilde{\sigma}_r^2$ closer to the true σ_r^2.

The free signal comes from news, rumors, and any source other than Market Insight, and it generates randomness in product entry patterns.[25] The free signal is assumed to be the true r_{jt} plus an iid normal distributed noise $\nu_{jkt,0}$:

$$\theta_{jkt,0} = r_{jt} + \nu_{jkt,0},\ \nu_{jkt,0} \sim \mathcal{N}(0, \sigma_k^2) \tag{4.8}$$

Sellers differ in their free signal variance σ_k^2 because some are better-informed than others. They are either more experienced with finding fashion trends or better at analyzing the competition.

When subscribing to Market Insight, the seller receives the additional signal about r_{jt} from observing sales by other sellers. The additional signal $\theta_{jkt,MI}$ is also r_{jt} plus noise $\nu_{jkt,MI}$:

$$\theta_{jkt,MI} = r_{jt} + \nu_{jkt,MI},\ \nu_{jkt,MI} \sim \mathcal{N}(0, \sigma_{MI}^2) \tag{4.9}$$

The variance of the Market Insight signal σ_{MI}^2 depends on the version of Market Insight,

[25] Without this free signal, a seller has the same prior for all products and will carry either all products or no product. For details, see the next section.

here assumed to be proportionate to the inverse of the number of sellers observed:[26]

$$\sigma_{MI}^2 = \sigma_m^2 / n_{MI} \tag{4.10}$$

n_{MI} is the number of observable sellers provided by the Market Insight subscription. For example, Market Insight pro has $n_{MI} = 30$ before the policy change and 60 after.

According to the Bayesian rule, seller k's posterior R_{jkt} of r_{jt} is given by

$$R_{jkt} = \mu_r + \frac{1/\sigma_k^2}{1/\tilde{\sigma}_r^2 + 1/\sigma_k^2 + n_{MI}/\sigma_m^2}(r_{jt} + \nu_{jkt,0} - \mu_r) + \frac{n_{MI}/\sigma_m^2}{1/\tilde{\sigma}_r^2 + 1/\sigma_k^2 + n_{MI}/\sigma_m^2}(r_{jt} + \nu_{jkt,MI} - \mu_r) \tag{4.11}$$

with a little abuse of notation: $n_{MI} = 0$ if the seller doesn't subscribe to Market Insight. The variance in this posterior is given by

$$\sigma_R^2 = \frac{1}{1/\tilde{\sigma}_r^2 + 1/\sigma_k^2 + n_{MI}/\sigma_m^2} \tag{4.12}$$

To simplify notation, a free signal $\nu_{jkt,0}$ with variance σ_k^2 and a Market Insight signal $\nu_{jkt,MI}$ with variance σ_m^2/n_{MI} is equivalent to the aggregate signal $\theta_{jkt,1} = \frac{1/\sigma_k^2}{1/\sigma_k^2 + n_{MI}/\sigma_m^2}\theta_{jkt,0} + \frac{n_{MI}/\sigma_m^2}{1/\sigma_k^2 + n_{MI}/\sigma_m^2}\theta_{jkt,MI}$ with variance $\sigma_\nu^2 = \frac{1}{1/\sigma_k^2 + n_{MI}/\sigma_m^2}$.

4.4.4 Assortment Decision

A seller's product assortment decision depends on the posterior belief about product profitabilities. For a tractable model of the assortment decision, we make the following assumptions:

A1. Monopolistic competition: Taobao sellers are so small that they don't think about how their product assortment will affect product decisions of rival sellers.

A2. Risk neutrality: Taobao sellers are risk-neutral.

[26] The underlying assumption is that observing one seller gives one independent signal of variance σ_m^2 about each lipstick. Therefore, observing n sellers gives n independent signals. Combined, n independent signals, each with variance σ_m^2, is equivalent to one signal with variance σ_m^2/n, according to the Bayesian rule.

A3. Zero sunk cost: Sellers don't incur any sunk cost in introducing new products.

A4. Timing: When deciding whether to carry product j, seller k will treat shocks in demand and marginal cost (u_{jkt}, v_{jkt}) as random.

The monopolistic competition assumption avoids strategic behaviors such as one seller carrying a particular item to prevent other sellers from carrying it. This is reasonable because among thousands of Taobao sellers, each would command a very small market share (with the largest seller commanding less than 1%). Risk neutrality guarantees that the assortment decision is product-by-product.[27] Otherwise, each seller will make decisions on the bundle of products to offer. With hundreds of lipstick brands on Taobao, the number of product bundles is too large to be handled by today's computer; risk neutrality is commonly assumed for firms. Zero sunk cost further simplifies the product assortment to a static decision, because carrying product j today doesn't have any impact on the payoff of carrying it tomorrow—it neither contains any information value (see previous section) nor incurs a sunk cost. Zero sunk cost makes sense in an online setting because the platform doesn't charge a seller to carry additional products; shelf space is essentially unlimited, and the promotion cost here is either a fixed cost (advertising) or a marginal cost (PPC).

With these simplifying assumptions, a seller carries a product if and only if the expected profit is larger than the fixed cost:

$$I_{jkt} = \mathbb{1}(log(\mathbb{E}(\pi_{jkt}|\mathcal{I}_{jkt}, \tilde{X}_{kt})) \geq \lambda_{jkt}) \tag{4.13}$$

where I_{jkt} is the assortment decision, an indicator of whether seller k carries product j in month t. λ_{jkt} is log(fixed cost).

The expected profit is conditional on seller characteristics $\tilde{X}_{kt}$ and the information set $\mathcal{I}_{jkt}$. $\tilde{X}_{kt}$ contains observed seller characteristics X_{kt} and unobserved seller cost heterogeneity c_k. The information set $\mathcal{I}_{jkt}$ contains posterior R_{jkt} and its variance σ_R^2 as well as

[27] A risk-averse seller has the incentive to carry more products to hedge the risk of any single product.

the variance of u and v:

$$log(\mathbb{E}(\pi_{jkt}|\mathcal{I}_{jkt}, \tilde{X}_{kt})) = C + R_{jkt} + (\beta_d + (1-\alpha)\beta_c)X_{kt} + (1-\alpha)c_k + \sigma_R^2/2 \qquad (4.14)$$

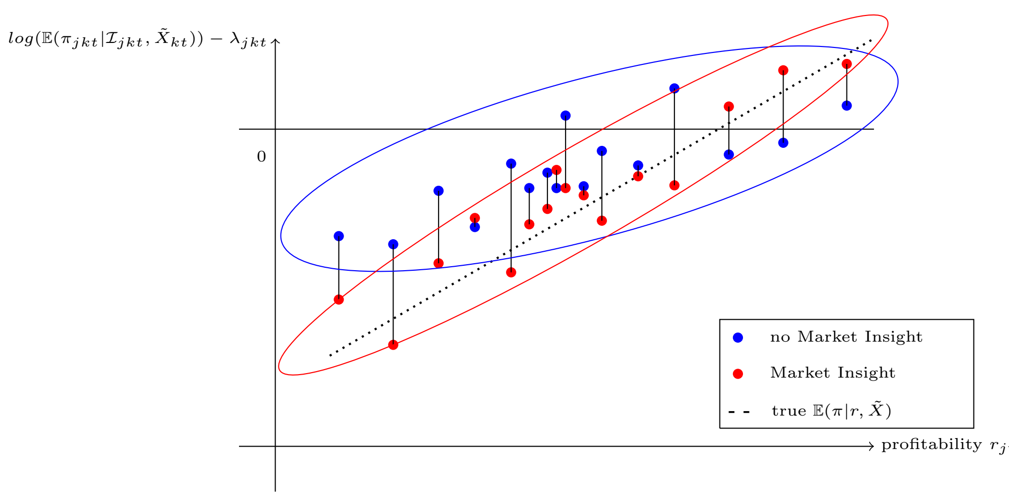
where

$$C \equiv log(M_t) + log(s_{0t}) + (1-\alpha)log(\frac{\alpha}{\alpha-1}) - log(\alpha) + \Sigma_{uv}/2, \text{ and}$$

$$\Sigma_{uv} \equiv \sigma_u^2 + (1-\alpha)^2\sigma_v^2 + 2(1-\alpha)\rho_{uv}$$

Σ_{uv} and σ_R^2 raise entry probability because for the seller who doesn't know u_{jkt}, v_{jkt}, or r_{jt} when making assortment decisions, larger variance means bigger chance of seeing a sizable profit.[28]

Figure 4.3: Seller Assortment Decision



- Y: indicator whether the seller will carry the product

To illustrate the assortment decision rule, we simulate the assortment decisions for a single seller in Figure 3. The y-axis is $log(\mathbb{E}(\pi_{jkt}|\mathcal{I}_{jkt}, \tilde{X}_{kt})) - \lambda_{jkt}$, and the x-axis is

[28]When $log(X) \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{E}(X) = exp(\mu + \sigma^2/2)$.

the profitability r_{jt}. Each pair of red and blue dots linked by a black line is a product. According to equation [13], the seller will carry a product if and only if $y \geq 0$. The blue and red dots represent seller assortment without and with Market Insight, respectively. The black dotted line is log expected profit (over u,v) minus log fixed cost conditional on true r_{jt}. The slope in red or blue dots represents how strongly the seller responds to r_{jt}. The Bayesian rule says that more precise information about r_{jt} means the seller will respond more to profitability r_{jt}; if the signals are random noise, then neither the posterior nor the assortment decision will respond to lipstick profitability. Therefore the two patterns predicted by the model are 1) because Market Insight gives the seller an additional signal of r_{jt}, the red dots have steeper slopes, and 2) the red dots are closer to the truth.

Assumptions on $(\sigma_k^2, \lambda_{jkt})$

To estimate the model, we impose further simplifying assumptions on the free signal variance σ_k^2 and the log(fixed cost) λ_{jkt}. I assume the free signal variance is the same within each Market Insight group {nonsubscriber, basic, pro}. The difference in σ_k^2 across Market Insight groups is important because it explains why some sellers subscribe to Market Insight: a seller needs Market Insight signals because the free signal is very noisy. Let $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$ denote σ_k^2 for {nonsubscriber, basic, pro} sellers, respectively.[29]

Regarding the log(fixed cost) λ_{jkt}, we assume it is seller-time specific and takes the following functional form:

$$\lambda_{kt} = \beta_\lambda \tilde{X}_{kt} + \tau_t + \lambda_m, \; m \in \{\text{nonsubscriber, basic, pro}\} \tag{4.15}$$

where λ_{kt} is affected by seller characteristics. For example, Gold sellers might need less promotion for its products, thus they will have lower fixed costs. τ_t captures variation in

[29] The assumption is that seller heterogeneous variance depends on current Market Insight choice. When we need to simulate seller choice of Market Insight, this assumption becomes very weird: regardless of Market Insight choice in the simulation, the seller always sticks to the original type. The main reason for this assumption to exist is to simplify computation. Furthermore, we will not simulate seller Market Insight choice for the later analysis. In this case, $\sigma_0^2, \sigma_1^2, \sigma_2^2$ can be interpreted as average within each seller group.

fixed cost over time.[30] I also allow a Market Insight group-specific component λ_m. This is the second source of seller heterogeneity that explains why sellers subscribe to Market Insight; let $(\lambda_0, \lambda_1, \lambda_2)$ denote λ_m for {nonsubscriber, MI basic, MI pro} sellers, respectively.

4.4.5 $\mathbb{P}(I_{jkt} = 1)$ and the Benefit of Market Insight to Sellers

This section documents the conditional probability that seller k carries product j in month t, $\mathbb{P}(I_{jkt} = 1|r_{jt}, \tilde{X}_{kt}, \sigma_k^2, m_{kt})$, which will be used for the MLE estimation. The probability is conditional on product j's profitability, the sellers's characteristics, her free signal variance and fixed cost, and the version of her Market Insight subscription. Then we derive the benefit of Market Insight to sellers.

The seller will carry the product if and only if $log(\mathbb{E}(\pi_{jkt}|\mathcal{I}_{jkt}, \tilde{X}_{kt})) \geq \lambda_{jkt}$, which happens when

$$\begin{aligned} w \cdot \nu_{jkt,1} &\geq -w \cdot (r_{jt} - \mu_r) - \sigma_R^2/2 - f(\tilde{X}_{kt}, \lambda) \\ \text{where } w =& \frac{1/\sigma_k^2 + n_{MI}/\sigma_m^2}{1/\tilde{\sigma}_r^2 + 1/\sigma_k^2 + n_{MI}/\sigma_m^2}, \\ \nu_{jkt,1} =& \theta_{jkt,1} - r_{jt}, \text{ and} \\ f(\tilde{X}_{kt}, \lambda) =& \mu_r + C + (\beta_d + (1-\alpha)\beta_c - \beta_{\lambda,X})X_{kt} + (1 - \alpha - \beta_{\lambda,c})c_k - \lambda_m - \tau_t \end{aligned} \tag{4.16}$$

where w is the Bayesian weight on the aggregate signal; it varies by who is the seller and whether the seller subscribes to Market Insight. $\nu_{jkt,1}$ is the noise, and $f(\tilde{X}_{kt}, \lambda)$ absorbs all the impacts of $\tilde{X}_{kt}$ on assortment through demand and marginal cost; it also absorbs the fixed cost.

With the aggregate signal noise being $\nu_{jkt,1} \sim \mathcal{N}(0, \sigma_\nu^2)$, the probability of seller k carrying product j at time t is given by

$$\mathbb{P}(I_{jkt} = 1|r_{jt}, \tilde{X}_{kt}, \sigma_k^2, m_{kt}) = \Phi(\frac{w \cdot (r_{jt} - \mu_r) + \sigma_R^2/2 + f(\tilde{X}_{kt}, \lambda)}{w \cdot \sigma_\nu}) \tag{4.17}$$

[30] Platform activities are frequent, such as its November shopping spree (like Amazon Prime Day); sellers will need extra hands and pay extra attention to inventory and logistics, and this increases fixed costs.

where σ_k^2 is seller heterogeneity in free signal variance, and $m_{kt} \in$ {nonsubscriber, basic, pro} indicates seller subscription to Market Insight. Both σ_k^2 and m_{kt} affect the seller information set $\mathcal{I}_{jkt}$, w, and σ_ν in equation [16].

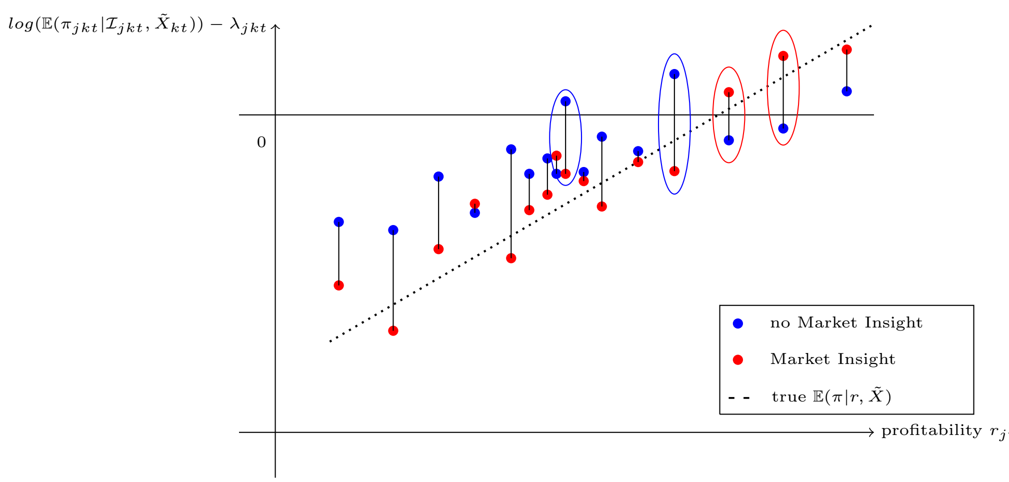
The benefit of subscribing to Market Insight is given by

$$B(m_{kt}, \tilde{X}_{kt}, \sigma_k^2) = \sum_j [\mathbb{P}(I_{jkt} = 1 | r_{jt}, \tilde{X}_{kt}, \sigma_k^2, m_{kt}) - \mathbb{P}(I_{jkt} = 1 | r_{jt}, \tilde{X}_{kt}, \sigma_k^2, \text{nonsubscriber})] \times \mathbb{E}(\pi_{jkt} | r_{jt}, \tilde{X}_{kt}) \tag{4.18}$$

For any product j, the benefit is the difference in probability of carrying product j multiplied by the expected profit (over (u, v)) conditional on realized r_{jt}.

The expected profit conditional on true r_{jt} isn't affected by Market Insight subscription. Market Insight only affects $\mathbb{P}(I_{jkt} = 1 | r_{jt}, \tilde{X}_{kt}, \sigma_k^2, m_{kt})$, the probability of carrying product j. When r_{jt} is large, Market Insight reminds the seller to carry it, raising $\mathbb{P}(I_{jkt} = 1)$; when r_{jt} is small, Market Insight is also helpful in letting the seller know it's unprofitable and leads to the seller dropping the product.

Figure 4.4: Benefit of Market Insight



- Y: indicator whether the seller will carry the product

To illustrate the benefit of Market Insight, we simulate seller assortment decisions with and without Market Insight in Figure 4. For products in red circles, without Market Insight, the posterior is relatively far from the truth, and the seller won't carry them. With Market Insight, the posteriors get closer to the truth, and sellers will carry them. For products in blue circles, Market Insight also gives the seller a better idea about the profit, but in this case makes the seller to drop these two products because the true profit is smaller than the fixed cost.

4.4.6 Seller Subscription to Market Insight

A seller may choose to subscribe to Market Insight and additionally choose between the basic and pro versions. Let the seller utility from choosing Market Insight be

$$\begin{cases} U_{mkt} = (-\alpha_m + \alpha_b B(m_{kt}, \tilde{X}_{kt}, \sigma_k^2) - p_{mt})/\sigma_\epsilon + \epsilon_{mkt} & , m \in \{\text{MI basic, MI pro}\} \\ U_{0kt} = 0 + \epsilon_{0kt} & , \text{nonsubscriber} \end{cases} \tag{4.19}$$

where U_{0kt} is the utility of not subscribing to Market Insight, and p_{mt} is the subscription fee that depends on Market Insight version. $B(m_{kt}, \tilde{X}_{kt}, \sigma_k^2)$ and p_{mt} are linear because both are in dollars. α_b reflects seller perception of Market Insight benefit; $\alpha_b = 1$ means the seller is fully aware of the benefit, while $\alpha_b < 1$ indicates the seller's perceived benefit is smaller than actual, which could be caused by a lack of Market Insight experience or lack of information. α_m captures the nonpecuniary cost of subscribing to Market Insight. It is in the model to fit the number of Market Insight subscribers. A large α_m could reflect that sellers need to hire data analysts to parse the Market Insight information or that many sellers aren't aware of Market Insight.[31] In addition, σ_ϵ accounts for the variance in perceived utility from Market Insight, because we normalize ϵ_{mkt} to be EV1 distributed with the standard deviation equal to 1.

[31]Market Insight is a relatively new data solution, created in August 2018.

4.5 Estimation

This section estimates demand and cost parameters $(d_{jt}, \gamma_{jt}, c_k, \alpha, \beta_d, \beta_c)$, assortment parameters $(\tilde{\sigma}_r^2, \sigma_m^2, \sigma_0^2, \sigma_1^2, \sigma_2^2, \beta_\lambda, \tau_t, \lambda_0, \lambda_1, \lambda_2)$, and Market Insight choice parameters $(\alpha_b, \alpha_m, \sigma_\epsilon)$. For demand estimation, there exist many zero market shares in the data because online sellers are quite small. we use the Markov Chain Monte Carlo (MCMC) method to overcome this problem. Then we estimate the assortment parameters, showing explicitly how the policy change helps the identification. Lastly, we estimate the Market Insight choice parameters.

4.5.1 Demand Estimation

Due to the EV1 distributed error term in equation [5], we get

$$log(q_{jkt}) = log(q_{0t}) + d_{jt} + \beta_d X_{kt} - \alpha log(p_{jkt}) + u_{jkt} \tag{4.20}$$

A problem in estimating this equation is that in many cases, $q_{jkt} = 0$; thus the left-hand side is not defined. Furthermore, the high-dimensional product demand d_{jt} makes it impossible to estimate this equation through maximum likelihood.

MCMC has been widely applied in cases of demand estimation with high-dimensional fixed effects, and it also deals with the zero-sales problem in this case. While the true data-generating process is given in equation [20], $LHS < 0$ is not possible in reality because it isn't possible to sell less than 1 unit; in this case, the observed $q_{jkt} = 0$. This means we need to draw the LHS when observed sales equal zero; MCMC with Gibbs sampling provides a systematic way of drawing the LHS. Having the step s-1 estimates of $(q_{0t}, d_{jt}, \beta_d, \alpha, \sigma_u)$, we can draw a new $log(q_{jkt})$ from $\mathcal{N}(log(q_{0t}) + d_{jt} + \beta_d X_{kt} - \alpha log(p_{jkt}), \sigma_u^2)$ but truncated below 0. For details of Gibbs sampling, see appendix B.3.

The remaining problem is price endogeneity. Recall that

$$log(p_{jkt}) = log(\frac{\alpha}{\alpha - 1}) + \underbrace{\gamma_{jt} + c_k + \beta_c X_{jkt} + v_{jkt}}_{log(c_{jkt})} \tag{4.21}$$

In equation [20], $log(p_{jkt})$ is correlated with u_{jkt} if u_{jkt} and v_{jkt} are correlated, which is expected. For example, if a seller pays a higher PPC price for a product, the v_{jkt} will be higher for that seller-product pair; this will also produce a higher search ranking and thus higher u_{jkt}.

The solution is to estimate equations [20] and [21] together. From equation [21] we get predicted $\hat{v}_{jkt}$; then in equation [20], $log(p_{jkt})$ is no longer correlated with u_{jkt}, conditional on $\hat{v}_{jkt}$. Essentially, this is using the seller cost heterogeneity c_k as the instrument for p_{jkt}. The instrument is valid because u_{jkt} is not correlated with any seller characteristics by assumption. On the other hand, seller cost c_k is correlated with price because it's a cost shifter.

For seller cost c_k to be a valid IV, the key assumption is that there is no unobserved seller heterogeneity in demand—there is no seller reputation that's observable by consumers but not by the platform and econometrician. This assumption isn't as stringent as it seems; many seller features—like Gold status—that attract consumers are actually created and maintained by Taobao.

Demand Estimation Results

Table 2 shows the demand and cost estimates. Estimated $\alpha = 4.5$, which translates into a markup $\frac{p-mc}{p} = 22\%$. This agrees with experiences of Taobao staff.

Table 4.2: Demand Estimates

	Demand coef..	Cost coef.
α	4.50	
Gold	0.62	0.03
Global	0.12	0.00
Stars	-0.02	-0.03
log(fans)	0.25	0.02

Note that Gold status has a large impact on consumer demand: it raises demand by 62%. This is consistent with the consensus notion that reputation plays a big role for online sellers (for example, see Hui et al. (2018)). Furthermore, Gold status also affects seller search rankings, thus boosting consumer demand even more. Number of fans also has a large impact: when the seller has twice as many fans (loyal customers), demand for any product rise by 25%. The coefficient on seller star level is very close to zero. Although higher star level indicates a higher number of past sales, it also means the seller is less sensitive to bad reviews. Therefore consumers have mixed feelings about star level.[32] None of the seller characteristics, however, has a discernible impact on marginal cost: β_c are all close to zero.

Furthermore, we am able to construct product profitability $r_{jt} \equiv d_{jt} - \alpha\gamma_{jt}$ from product demand and cost states (d_{jt}, γ_{jt}), with demand estimates. The r_{jt} is the key object for seller assortment decisions because it is the object that the seller learns from Market Insight. I also document some evidence of learning about r_{jt}: after the policy change, treated sellers begin carrying products with higher profitability r_{jt} (see appendix B.2).

[32]For new sellers, one bad review can alter star level drastically; for more established sellers, a bad review becomes diluted by all historical reviews and will have little impact. This means established sellers might not be concerned about bad reviews. Since bad reviews are one of the few tools consumers can use to hold sellers accountable, there is widespread distrust of high-star sellers, leading to the phrase, "High star sellers tend to take advantage of consumers" in China.

4.5.2 Estimating Assortment Parameters

Assortment parameters include the common prior variance $\tilde{\sigma}_r^2$, the Market Insight group-specific free signal variance $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$, the Market Insight signal variance σ_m^2 (of observing one seller), and the fixed-cost parameters $(\beta_\lambda, \tau_t, \lambda_0, \lambda_1, \lambda_2)$.

Since the conditional likelihood is written down in equation [17], we can estimate all assortment parameters using maximum likelihood, given the demand parameters as well as the product profitability r_{jt} from the demand estimation.

The policy change helps identify the assortment parameters. For example, the Bayesian model says that signal variances σ_m^2 (MI) and $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$ (free) determine how responsive the assortment decision is to profitability r_{jt}. Using the policy change, we can recover the causal effect of observing 30 more sellers on assortment responsiveness among treated sellers. The structure of the model then maps that causal effect to signal variances $(\sigma_m^2, \sigma_0^2, \sigma_1^2, \sigma_2^2)$.

To be specific, consider equation [17] with simplified notation:

$$\begin{aligned}
&\mathbb{P}(I_{jkt} = 1 | r_{jt}, \tilde{X}_{kt}, \sigma_k^2, m_{kt}) \\
&\qquad = \Phi\left(\frac{w \cdot (r_{jt} - \mu_r) + \sigma_R^2/2 + f(\tilde{X}_{kt}, \lambda)}{w \cdot \sigma_\nu}\right) \\
&\qquad = \Phi(\beta_0 + \beta_r \cdot (r_{jt} - \mu_r) + g(\tilde{X}_{kt}, \lambda))
\end{aligned}$$

where β_r is the responsiveness of the assortment decision to profitability r_{jt}, and β_0 is the intercept of the rate of carrying a product. The Bayesian learning model says responsiveness β_r is determined by free signal variances $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$ and MI signal variance σ_m^2: the more precise the aggregate signal, the higher the β_r (Figure 4). In particular, a treated (MI pro) seller's β_r is given by

$$\beta_r = \begin{cases} \sqrt{1/\sigma_2^2 + 30/\sigma_m^2} & \text{, Market Insight pro, before August 2019, observe 30} \\ \sqrt{1/\sigma_2^2 + 60/\sigma_m^2} & \text{, Market Insight pro, after August 2019, observe 60} \end{cases} \tag{4.22}$$

The pro seller's free signals have same precision before and after the policy change, so if the seller becomes more responsive to r_{jt} afterward, then it must stem from the more precise

Market Insight signals that follow the policy change.

Assortment Estimation Results

Table 3 displays the assortment parameters. The first panel shows the parameters that determine seller responsiveness to profitability r_{jt}, β_r. They are the free signal variances $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$ and the Market Insight signal variance σ_m^2.

To get a sense of the magnitude of signal variance, compare it to the prior. For nonsubscribers and basic subscribers, their free signals are approximately one-fifth as precise as the common prior. This reflects the data pattern that these sellers' product choices don't correspond strongly to profitability.

Pro subscribers have even noisier free signals; their free signals are less than one-tenth as precise as the prior. This reflects the data pattern where before the policy change, even though pro subscribers have 30 sellers to observe, they are still less responsive to profitability r_{jt} than other sellers (for details, see Appendix B.2, evidence of learning). Such noisy free signals means without Market Insight, current pro sellers cannot distinguish which products are profitable, explaining why they have chosen be pro subscribers.

Table 4.3: Assortment Estimates

	Coefficient
$\mathbb{V}$(free signal), nonsubscriber sellers	31
$\mathbb{V}$(free signal), MI basic sellers	27
$\mathbb{V}$(free signal), MI pro sellers	81
$\mathbb{V}$(MI signal), obs. one seller	1575
log(fixed cost), nonsubscriber sellers fixed effect λ_0	7.8
log(fixed cost), basic sellers fixed effect λ_1	8.0
log(fixed cost), pro sellers fixed effect λ_2	8.0
$\mathbb{V}$(prior of profitability r_{jt})	5.75

The second panel shows the parameters (fixed costs $(\lambda_0, \lambda_1, \lambda_2)$ and prior variance $\tilde{\sigma}_r^2$) that fit the level of the probability of carrying each lipstick brand. The variance in prior of profitability r_{jt}, $\tilde{\sigma}_r^2$ is larger than the true variance in r_{jt}, which is 1.30, reflecting that sellers are not well informed.

In addition, the estimated fixed costs are relatively large. For example, the average fixed cost for nonsubscribers is 3,400 USD. Since the fixed costs are there to fit the level of the probability of carrying any product, the large estimated fixed costs result from the data pattern that each seller only carries a narrow selection of products. One interpretation is that λ, the log(fixed cost), has a high variance. In this case, sellers stay out of many products, not because the fixed cost is so large, but because the realized fixed cost can be very large.[33] Alternatively, it could be interpreted as the mental cost of carrying one more brand being high. The mental cost would be the cost of managing many brands or of learning about a new brand.

Comparison with Reduced-Form Findings

With the assortment estimates, we can simulate seller product choices before and after the policy change among the treated sellers and see how seller revenue and number of lipsticks carried respond to this change. The model predicts that the treated sellers will raise their rate of carrying a brand by 10%, and their monthly revenue will rise by 26%.

Recall the reduced-form finding that the rate of introducing a new product goes up by 20%, and the monthly revenue goes up by 40% after the policy change. The model prediction and the reduced-form findings are compatible with each other, suggesting the model is doing a good job capturing the relationship between seller assortment decisions and Market Insight information. However, the model prediction is smaller in magnitude than the reduced-form findings, partly because all modelled gains are at the brand level—Market Insight tells a seller one brand is more profitable than another. In reality, the information in Market Insight is even finer, revealing the profitability of, for example, Dior 999 (Dior pure red). Therefore, it's expected that the model predicts a lower bound of the treatment

[33] The model is compatible with variance in fixed cost. By assuming sellers don't know the realization of fixed cost when making the assortment decision, λ_{kt} in equation [15] is true log(fixed cost) $\tilde{\lambda}_{kt} + \sigma_{\lambda}^2/2$.

effects.[34]

4.5.3 Estimating Market Insight Choice Parameters

To recover Market Insight choice parameters $(\alpha_b, \alpha_m, \sigma_\epsilon)$, we estimate a logit (seller) demand for Market Insight. The key parameter is α_b, how sellers' Market Insight choices depend on the predicted benefit of Market Insight. The estimated α_b is 0.0015, reflecting the model did not do a good job capturing seller Market Insight choice (Table 4).

Table 4.4: Market Insight Choice Estimates

	Coefficient
perception of Market Insight benefit α_b	0.0015
non-pecuniary cost of subscribing to Market Insight α_m	996
std of perceived Market Insight utility σ_ϵ	378

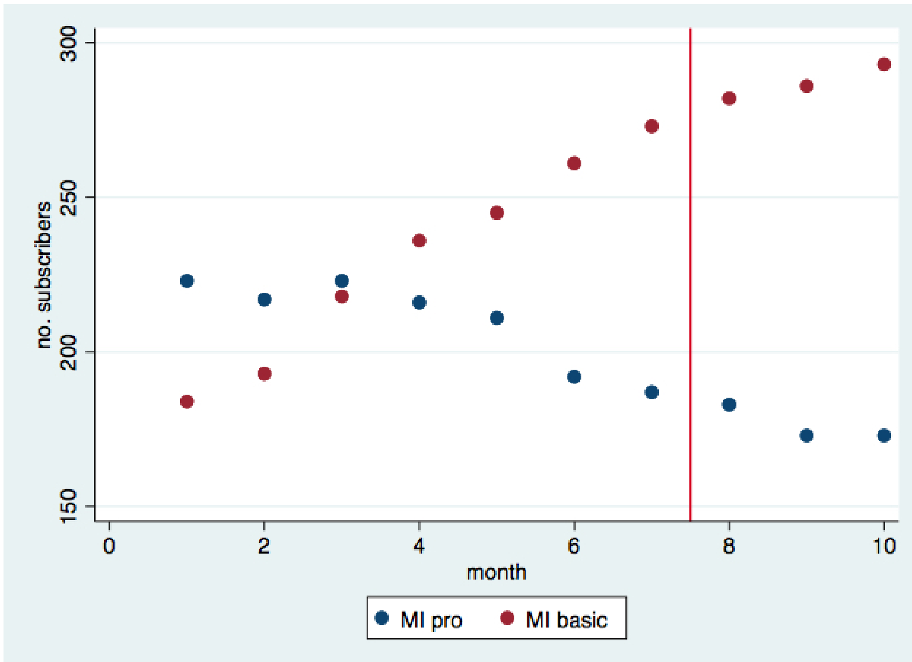
The model's inability to explain seller Market Insight choice likely has little to do with modeling; the main reason is too much noise regarding Market Insight choice in the data. Recall that the policy change leads to a 40% increase in seller revenue, which means the benefit of Market Insight pro increases significantly after the policy change regardless of model specifications. However, this increase in benefit doesn't lead to more sellers subscribing to Market Insight pro in the data. As shown in Figure 5, the number of pro sellers continues to fall throughout the sample period, while the number of basic sellers keeps rising. This shows that the benefit of Market Insight, no matter the model specification, is going to be poorly correlated with sellers' Market Insight choices; there is just too much noise in the Market Insight decision.[35] Due to the large amount of noise in

[34]The reason model estimates are only possible at the brand level is due to too much noise in the data about lipstick color and texture.

[35]One source of noise was the creation of Market Insight itself in August 2018. To promote this new data tool, any seller subscribing to its substantially less informative predecessor, "Market

seller selection of Market Insight, we restrict the use of it in the following counterfactual analysis.

Figure 4.5: Number of Market Insight Basic and Pro Sellers over Time



4.6 Counterfactual

With the demand and assortment estimates, we can answer two questions: 1) What is the value of Market Insight to consumers and the platform? and 2) What is the optimal design of Market Insight for consumers and the platform?

Comprehension" (also offered in basic and pro), automatically gets upgraded to the new Market Insight. Users retain their existing basic or pro subscription levels and expiration dates. Because the Market Comprehension subscriptions were for at least one year, those upgraded subscriptions gradually expire following August 2018. Market Insight pro is far more expensive than Market Comprehension pro, hence many upgraded subscribers will fail to renew at the pro level when their old subscription expires, in many cases switching to the less expensive basic level. This explains the decreasing trend over time in the number of pro subscribers while basic subscribers increase in number.

4.6.1 Value of Market Insight

To determine what value Market Insight provides to consumers and to the platform, we simulate what products sellers will offer and at what price. we then compute total platform income as well as consumer welfare.

I assume platform income is proportionate to total sales on the platform because sales is the single-most widely used measure that Taobao employs to show growth and present the competition landscape with other platforms. Also, Taobao's primary source of income is seller advertising, and for other online e-commerce platforms, advertising income is proportionate to total sales.[36]

Absent Market Insight, platform total sales drop by 8.1%, and consumer welfare drops by 0.70%.[37] The reason for this is that when current Market Insight sellers lose Market Insight information, they reduce their number of products by 22%. This reduces product variety offered on the platform as a whole and leads to lower total sales and lower consumer welfare.

4.6.2 The Tradeoff

Is it always better for the platform to provide more information? Not necessarily. The tradeoff here is between what lipsticks are offered versus how many lipsticks are offered. When sellers gain awareness of products that are more profitable, they tend to carry those products. Since these products are also those with higher demand, lower cost, and more revenue-generating capacity, offering more of them benefits consumers as well as the platform. On the other hand, giving away too much information bears the risk that sellers only sell the most profitable products, thereby reducing variety; fewer products offered reduces

[36] I have to admit this is my speculation based on readings of similar platforms such as Amazon. we dont know if this is the case for Taobao.

[37] The magnitude of the consumer welfare reduction mainly depends on s_0, the market share of outside options. Here we define s_0 to be $1 - \frac{\text{no. transactions}}{\text{no. of people who click on at least one lipstick}}$. In this case, $s_0 = 0.86$.

benefits for consumers and the platform.[38]

Table 4.5: Tradeoff in Providing More Information

	I	II	III
consumer welfare	+0.79%	+1.45%	+1.37%
total lipstick sales	+9.83%	+17.44%	+16.42%
no. seller-prod. pairs	−2.4%	−8.2%	−10.2%
mean r_{jt}	+0.01	+0.05	+0.06
no. sellers to observe	30	180	360

- compare to no Market Insight
- keep the Market Insight sellers fixed
- equilibrium: seller respond to change in outside option s_0 by carrying more or less products
- quality improvement measured in std

To show this tradeoff, we simulate the product assortment decisions under different information policies in Table 5. Column 1 displays simulated consumer welfare and total platform sales if current Market Insight subscribers (both basic and pro) have 30 sellers to observe, compared to the benchmark where no one has Market Insight. In columns 2 and 3, current subscribers are offered 180 and 360 sellers to follow, respectively. The tradeoff is illustrated in the second panel: as more information is revealed, variety goes down as informed sellers drop less profitable products. On the other hand, the lipsticks they keep (or add) are more profitable, driving up the mean profitability of lipsticks offered on the platform. The two forces compete with each other, so that the medium amount of information among the three cases promotes the highest consumer welfare as well as total platform sales.

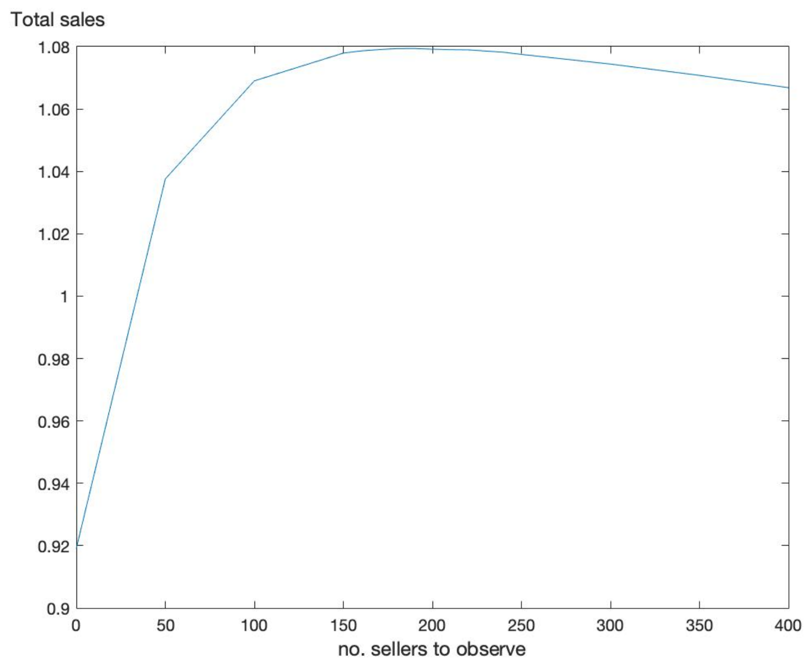
4.6.3 Platform-Optimal versus Socially Optimal Design

What is the platform-optimal information policy that maximizes total sales? To solve for this policy, we simulate total platform sales under different information policies (Figure 6).

[38]The Market Insight team at Taobao confirms that this is main tradeoff they face when adjusting Market Insight design. When they give subscribers more information, they see these sellers start to sell products similar to each other.

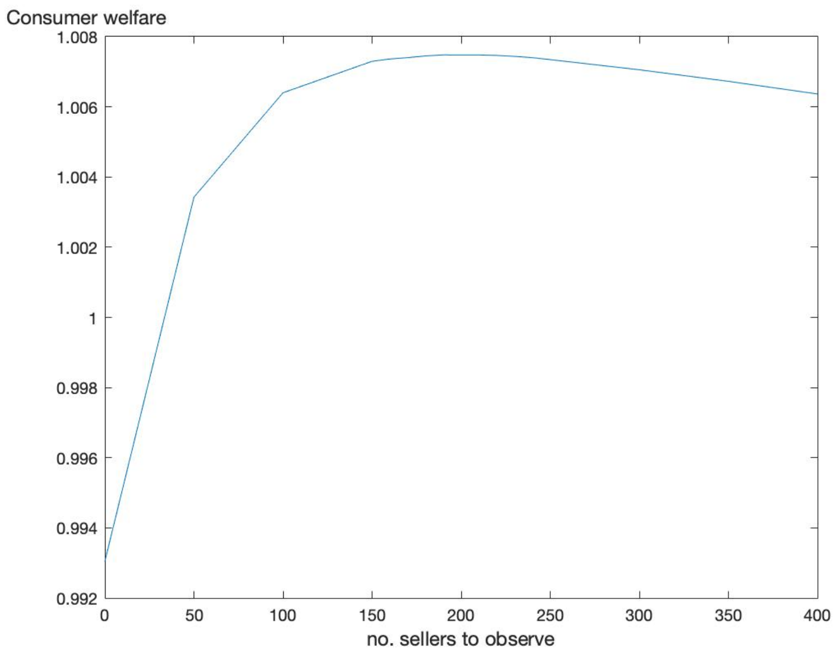
Total sales first increases and then decreases as subscribers are able to observe more sellers, reflecting the tradeoff highlighted in section 6.2. From Figure 6, the platform-optimal policy would be to let subscribers observe 185 sellers, which will increase total sales by another 8%.

Figure 4.6: Effect of Information Design on Platform Sales



y-axis: compare to current Market Insight design. For example, if new design allows 0 sellers to observe, then it is the same as section 6.1 with no Market Insight. In this case, the platform total sales goes down by 8%. For this graph, we fix Market Insight subscribers to be the current ones. I.e., no switcher in subscription is allowed.

Figure 4.7: Effect of Information Design on Consumer Welfare



How does the platform-optimal policy compare with a socially optimal policy that maximizes consumer welfare? Figure 7 shows that, like platform total sales, consumer welfare also first increases and then decreases with the amount of information provided.

[Figure 7 around here]

In fact, the platform-optimal policy isnt that different from the socially optimal policy: to maximize consumer welfare, Market Insight should let subscribers observe 200 sellers. To see how similar the platform-optimal policy is to the socially optimal policy, see Table 6. The first row lists the platform-optimal information policy and the resulting platform sales and consumer welfare compared to those under the current Market Insight policy. The second row corresponds to the socially optimal policy. We can see that the two policies produce very similar total sales and consumer welfare numbers.

Table 4.6: Platform vs Social Optimal

	no. sellers to observe	total sales	consumer welfare
max platform total sales	185	+7.94%	+0.745%
max consumer welfare	200	+7.92%	+0.749%

- benchmark: current Market Insight policy

The two optimal policies are very near to each other because platform preferences are very similar to those of consumers. Platform sales $\propto \frac{\sum_{jkt} exp(\delta_{jkt}+log(p_{jkt}))}{1+\sum_{jkt} exp(\delta_{jkt})}$ and consumer welfare $= log(1 + \sum_{jkt} exp(\delta_{jkt})) + ec.$[39] Note that without $log(p_{jkt})$, both consumers and the platform maximize $\sum_{jkt} exp(\delta_{jkt})$! Furthermore, price changes very little when the information policy changes. When the number of sellers to observe rises from 0 to 400, the weighted average price only varies within 0.2 USD. Since price doesn't change appreciably, it is no surprise that the platform and consumers share nearly identical preferences. In other words, in this case the platform behaves like a benevolent social planner that maximizes consumer welfare.

In this case, platform preference coincides with consumer preference because price doesn't vary a lot. In other words, here Market Insight has little effect in inducing a seller to sell lower cost lipsticks. Suppose in another sector, the benefit of Market Insight comes from helping sellers to find lower cost products to sell, then more information means lower overall price on the platform, which the platform doesn't necessarily like. In this case, consumer preference and platform preference can diverge.

4.7 Conclusion

Technological progress in data collection, storage, and analysis has made data on what sales look like more readily available. While firms rely more and more on these data to make decisions, there is little work on how these data change their decision-making, and how it affects consumers.

[39] δ_{jkt} is the mean consumer utility from product j, seller k, month t: consumer utility $U_{ijkt} = \delta_{jkt} + \epsilon_{ijkt}$); ec is the Euler constant.

This paper studies the impact of market intelligence data on online seller performance, product choice, and market outcomes by exploiting a unique setting where Taobao provides a market intelligence tool, Market Insight, to its sellers.

We document evidence that Market Insight significantly increases online lipstick seller revenue by influencing which lipstick brands and models they carry. In particular, it helps sellers find more profitable lipsticks to carry, which is a difficult problem in the lipstick sector given changing fashion trends, and hundreds of lipstick varieties to choose from.

Furthermore, through influencing online seller product choice, Market Insight also benefits the online platform and its consumers. Specifically, Market Insight leads to 8% higher total platform sales and 0.8% higher consumer welfare.

These results suggest that information policy can have significant impacts for e-commerce platforms, that is, aside from pricing strategy, the focus of the two-sided market literature of the past two decades. I show that by allowing subscribers to follow 185 other sellers, the e-commerce platform could further increase sales by 8%. It is also interesting to note that in this case, the platform has very similar preferences to that of a social planner who cares about consumer welfare; the optimal information policy for the platform is very similar to a socially optimal policy.

Note that the information design problem in this setting is not trivial: it is not true that the more information provided, the better off everyone will be. The key is the difference in incentives among sellers, the platform, and consumers. When sellers are more aware of market trends, they drop unprofitable products, which could reduce product variety on the platform. This could harm consumers, lower their chance of buying on the platform, and therefore could also harm the platform.

For future research, one interesting direction is the process of data creation. This paper shows that providing more information could reduce product variety on the platform. Given that the data also comes from the products offered on the platform, less variety means less data. Considering not only how data is used but also how it is created gives a more complete answer to the information design question. Another interesting direction is the learning network. With additional data on who learns from whom, one could gain further insight into the social learning process.

Chapter 5

Conclusion

In this dissertation, I study the impact of the recent trends in the retail industry. In chapter 2, I describe the data collection and analysis process in the offline and online retail markets. In chapter 3, I study the impact of store brands on consumer welfare by carefully characterizing the effect of store brands on other products' prices. In chapter 4, I study the impact of data and information on online seller's decision-making and market outcomes, exploiting a unique setting in which an e-commerce platform provides its sellers with a market intelligence tool called Market Insight. First I find Market Insight helps online sellers choose better products and increases their sales. Second I show the current design of Market Insight benefits consumers and the platform, though providing "too much" information through Market Insight could be harmful. Finally I solve for the platform-optimal design of Market Insight and show that both the platform and consumers benefit from this design. I also compare the platform-optimal design to the socially optimal one., and find they are very close to each other; i.e., in this case, the platform acts like a benevolent social planner.

Appendix A

Appendix to Chapter 3

A.1 (p_1, p_2)

$$\begin{cases} \frac{\partial s_1}{\partial p_1} = -\alpha(s_1 - s_1^2) \\ \frac{\partial s_2}{\partial p_2} = -\alpha(s_2 - s_2^2) \end{cases}$$

and

$$\frac{\partial s_2}{\partial p_2} = \alpha s_1 s_2$$

A.2 F_{δ_1}

$$\begin{aligned} F_{\delta_1} &= -\frac{1}{\alpha}\frac{\partial}{\partial \delta_1}[exp(\delta_1 - \alpha p_1) + exp(\delta_2 - \alpha p_2)] \\ &= -\frac{1}{\alpha} exp(\delta_1 - \alpha p_1)(1 - \alpha \frac{\partial p_1}{\partial \delta_1}) \\ &= -\frac{1}{\alpha} \frac{exp(\delta_1 - \alpha p_1)}{1 + exp(\delta_1 - \alpha p_1)} \end{aligned}$$

to derive $\frac{\partial p_1}{\partial \delta_1}$

$$\begin{aligned} p_1 &= c_1 + \frac{1}{\alpha}(1 + exp(\delta_1 - \alpha p_1) + exp(\delta_2 - \alpha p_2)) \\ \Rightarrow \frac{\partial p_1}{\partial \delta_1} &= \frac{1}{\alpha} exp(\delta_1 - \alpha p_1)(1 - \alpha \frac{\partial p_1}{\partial \delta_1}) \\ \Rightarrow \frac{\partial p_1}{\partial \delta_1} &= \frac{1}{\alpha} \frac{exp(\delta_1 - \alpha p_1)}{1 + exp(\delta_1 - \alpha p_1)} \end{aligned}$$

A.3 F_{p_2}

$$
\begin{aligned}
F_{p_2} &= 1 - \frac{1}{\alpha}\frac{\partial}{\partial p_2}[exp(\delta_1 - \alpha p_1) + exp(\delta_2 - \alpha p_2)] \\
&= 1 - \frac{1}{\alpha}[exp(\delta_1 - \alpha p_1)(-\alpha)\frac{\partial p_1}{\partial p_2} + exp(\delta_2 - \alpha p_2)(-\alpha)] \\
&= 1 + [exp(\delta_1 - \alpha p_1)\frac{\partial p_1}{\partial p_2} + exp(\delta_2 - \alpha p_2)] \\
&= 1 + \frac{exp(\delta_2 - \alpha p_2)}{1 + exp(\delta_1 - \alpha p_1)}
\end{aligned}
$$

to derive $\frac{\partial p_1}{\partial p_2}$

$$
\begin{aligned}
p_1 &= c_1 + \frac{1}{\alpha}(1 + exp(\delta_1 - \alpha p_1) + exp(\delta_2 - \alpha p_2)) \\
\Rightarrow \frac{\partial p_1}{\partial p_2} &= \frac{1}{\alpha}[exp(\delta_1 - \alpha p_1)(-\alpha\frac{\partial p_1}{\partial p_2}) + exp(\delta_2 - \alpha p_2)(-\alpha)] \\
\Rightarrow \frac{\partial p_1}{\partial p_2} &= -\frac{exp(\delta_2 - \alpha p_2)}{1 + exp(\delta_1 - \alpha p_1)}
\end{aligned}
$$

A.4 $\frac{\partial w}{\partial \delta_1}$, some preparation

- from equation [4], $p_1 - c_1 = p_2 - w = \frac{D}{\alpha}$; $\Pi_r = \frac{D}{\alpha}(s_1 + s_2) = \frac{1}{\alpha}(N_1 + N_2)$, and $D_r = \frac{D}{\alpha}\frac{N_1}{1+N_1}$

- Since

$$
\begin{cases}
p_1 - c_1 = \frac{1}{\alpha}(1 + N_1 + N_2) \\
p_2 - w = \frac{1}{\alpha}(1 + N_1 + N_2)
\end{cases}
$$

taking derivative wrt δ_1, and keep in mind that $N_1 = exp(\delta_1 - \alpha p_1)$ and $N_2 = exp(\delta_2 - \alpha p_2)$, we get

$$
\begin{cases}
\frac{\partial p_1}{\partial \delta_1} = \frac{1}{\alpha}(N_1(1 - \alpha\frac{\partial p_1}{\partial \delta_1}) + N_2(-\alpha)\frac{\partial p_2}{\partial \delta_1}) \\
\frac{\partial p_2}{\partial \delta_1} = \frac{\partial p_1}{\partial \delta_1}
\end{cases}
$$

solving for $(\frac{\partial p_1}{\partial \delta_1}, \frac{\partial p_2}{\partial \delta_1})$, we get

$$
\frac{\partial p_1}{\partial \delta_1} = \frac{\partial p_2}{\partial \delta_1} = \frac{1}{\alpha}s_1
$$

- furthermore,

$$\begin{cases} \frac{\partial N_1}{\partial \delta_1} = N_1(1 - s_1) \\ \frac{\partial N_2}{\partial \delta_1} = -N_2 s_1 \\ \frac{\partial D}{\partial \delta_1} = \frac{\partial N_1}{\partial \delta_1} + \frac{\partial N_2}{\partial \delta_1} = s_1 \end{cases}$$

A.5 G_{δ_1}

$$
\begin{aligned}
G_{\delta_1} &= \frac{1}{\alpha}\frac{\partial}{\partial\delta_1}[N_1 + N_2 - D\frac{N_1}{1+N_1}] - (w-c_2)\frac{\partial}{\partial\delta_1}\frac{N_2}{D} \\
&= \frac{1}{\alpha}[\frac{\partial D}{\partial\delta_1} - \frac{\partial D}{\partial\delta_1}\frac{N_1}{1+N_1} - D\frac{\partial}{\partial\delta_1}\frac{N_1}{1+N_1}] - \frac{w-c_2}{D^2}(\frac{\partial N_2}{\partial\delta_1}D - N_2\frac{\partial D}{\partial\delta_1}) \\
&= \frac{1}{\alpha}[\frac{\partial D}{\partial\delta_1}\frac{1}{1+N_1} - D\frac{\frac{\partial N_1}{\partial\delta_1}(1+N_1) - N_1\frac{\partial N_1}{\partial\delta_1}}{(1+N_1)^2}] - \frac{w-c_2}{D^2}(-N_2 s_1 D - N_2 s_1) \\
&= \frac{1}{\alpha}[\frac{\partial D}{\partial\delta_1}\frac{1}{1+N_1} - \frac{D\frac{\partial N_1}{\partial\delta_1}}{(1+N_1)^2}] + (w-c_2)s_2\frac{s_1 D + s_1}{D} \\
&= \frac{1}{\alpha}[\frac{s_1}{1+N_1} - \frac{DN_1(1-s_1)}{(1+N_1)^2}] + \Pi_m s_1(1+s_0) \\
&= \frac{1}{\alpha}[\frac{s_1}{1+N_1} - \frac{DN_1(1-s_1)}{(1+N_1)^2}] + (\Pi_r - D_r)s_1(1+s_0) \\
&= \frac{1}{\alpha}[\frac{s_1}{1+N_1} - \frac{DN_1(1-s_1)}{(1+N_1)^2}] + (\frac{1-s_0}{\alpha s_0} - \frac{N_1}{1+N_1}\frac{1}{\alpha s_0})s_1(1+s_0) \\
&= \frac{1}{\alpha}[\frac{s_1}{1+N_1} - \frac{DN_1(1-s_1)}{(1+N_1)^2}] + (\frac{1}{1+N_1}\frac{1}{\alpha s_0} - \frac{1}{\alpha})s_1(1+s_0) \\
&= \frac{1}{\alpha}[\frac{s_1}{1+N_1} - \frac{DN_1(1-s_1)}{(1+N_1)^2}] + \frac{1}{\alpha}(\frac{D}{1+N_1} - 1)s_1(1+s_0) \\
&= \frac{1}{\alpha}[\frac{s_1}{1+N_1} - \frac{DN_1(1-s_1)}{(1+N_1)^2}] + \frac{1}{\alpha}\frac{N_2}{1+N_1}s_1(1+s_0) \\
&= \frac{1}{\alpha}[\frac{s_1}{1+N_1} - \frac{DN_1(1-s_1)}{(1+N_1)^2} + \frac{N_2}{1+N_1}s_1(1+s_0)] \\
&= \frac{1}{\alpha(1+N_1)^2}[s_1(1+N_1) - DN_1(1-s_1) + N_2(1+N_1)s_1(1+s_0)] \\
&= \frac{D^2}{\alpha(1+N_1)^2}[s_1(s_0+s_1)s_0 - s_1(1-s_1) + s_2(s_0+s_1)s_1(1+s_0)] \\
&= \frac{D^2}{\alpha(1+N_1)^2}[s_1(1-s_2)s_0 - s_1(1-s_1) + s_2(1-s_2)s_1(1+s_0)] \\
&= \frac{D^2}{\alpha(1+N_1)^2}[s_1 s_0 - s_0 s_1 s_2 - s_1 + s_1^2 + s_2 s_1 - s_2^2 s_1 + s_2 s_1 s_0 - s_2^2 s_1 s_0] \\
&= \frac{D^2}{\alpha(1+N_1)^2}[s_1 s_0 - s_1 + s_1^2 + s_2 s_1 - s_2^2 s_1 - s_2^2 s_1 s_0] \\
&= \frac{D^2}{\alpha(1+N_1)^2}[s_1 s_0 - s_1(1 - s_1 - s_2) - s_2^2 s_1 - s_2^2 s_1 s_0] \\
&= -\frac{D^2}{\alpha(1+N_1)^2}s_2^2 s_1(1+s_0)
\end{aligned}
$$

A.6 H_{δ_1}, some preparation

- since

$$p_1 - c_1 = \frac{1}{\alpha}(1 + N_1 + N_2)$$

taking derivative wrt δ_1, we get

$$\frac{\partial p_1}{\partial \delta_1} = \frac{1}{\alpha} N_1 (1 - \alpha \frac{\partial p_1}{\partial \delta_1}) = \frac{1}{\alpha} \frac{N_1}{N_1 + 1}$$

- and

$$\frac{\partial N_1}{\partial \delta_1} = \frac{\partial D}{\partial \delta_1} = \frac{N_1}{N_1 + 1}$$

- furthermore, since

$$p_2 - w = p_1 - c_1$$

we have

$$\frac{\partial w}{\partial \delta_1} = -\frac{\partial p_1}{\partial \delta_1} = -\frac{1}{\alpha} \frac{N_1}{N_1 + 1}$$

A.7 Cannot sign H_{δ_1}

$$
\begin{aligned}
H_{\delta_1} = &\frac{1}{\alpha}\frac{\partial}{\partial\delta_1}[N_1 + N_2 - D\frac{N_1}{1+N_1}] - N_2\frac{\partial}{\partial\delta_1}\frac{w-c_2}{D}\\
=&\frac{1}{\alpha}[\frac{\partial N_1}{\partial\delta_1} - (\frac{\partial D}{\partial\delta_1}\frac{N_1}{1+N_1} + \frac{D}{(1+N_1)^2}(\frac{\partial N_1}{\partial\delta_1}(1+N_1) - N_1\frac{\partial N_1}{\partial\delta_1}))]\\
&-\frac{N_2}{D^2}(\frac{\partial w}{\partial\delta_1}D - (w-c_2)\frac{\partial D}{\partial\delta_1})\\
=&\frac{1}{\alpha}[\frac{\partial N_1}{\partial\delta_1}\frac{1}{1+N_1} - \frac{D}{(1+N_1)^2}\frac{\partial N_1}{\partial\delta_1}] - \frac{N_2}{D}\frac{\partial w}{\partial\delta_1} + \frac{(w-c_2)s_2}{D}\frac{\partial D}{\partial\delta_1}\\
=&\frac{1}{\alpha}\frac{\partial N_1}{\partial\delta_1}[\frac{1}{N_1+1} - \frac{D}{(1+N_1)^2} + \frac{N_2}{D} + \frac{N_2}{D(1+N_1)}]\\
=&\frac{N_1}{\alpha(1+N_1)^3D}[(1+N_1)D - D^2 + N_2(1+N_1)^2 + N_2(1+N_1)]\\
=&\frac{N_1}{\alpha(1+N_1)^3D}[-N_2D + N_2(1+N_1)^2 + N_2(1+N_1)]\\
=&\frac{N_1}{\alpha(1+N_1)^3D}[-N_2^2 + N_2(1+N_1)^2]\\
=&\frac{N_1N_2}{\alpha(1+N_1)^3D}[(1+N_1)^2 - N_2]
\end{aligned}
$$

A.8 H_{p_2}, some preparation

- since

$$p_1 - c_1 = \frac{1}{\alpha}(1 + N_1 + N_2)$$

taking derivative wrt p_2, we get

$$\frac{\partial p_1}{\partial p_2} = \frac{1}{\alpha}[N_1(-\alpha)\frac{\partial p_1}{\partial p_2} + N_2(-\alpha)] = -\frac{N_2}{N_1+1}$$

- and

$$
\begin{cases}
\frac{\partial N_1}{\partial p_2} = N_1(-\alpha)\frac{\partial p_1}{\partial p_2} = \alpha\frac{N_1N_2}{N_1+1}\\
\frac{\partial N_2}{\partial p_2} = N_2(-\alpha)\\
\frac{\partial D}{\partial p_2} = \frac{\partial N_1}{\partial p_2} + \frac{\partial N_2}{\partial p_2} = -\alpha\frac{N_2}{N_1+1}
\end{cases}
$$

- furthermore, since

$$p_2 - w = p_1 - c_1$$

we have

$$1 - \frac{\partial w}{\partial p_2} = \frac{\partial p_1}{\partial p_2}$$

$$\frac{\partial w}{\partial p_2} = 1 - \frac{\partial p_1}{\partial p_2} = \frac{D}{1+N_1}$$

A.9 $H_{p_2} < 0$

$$\begin{aligned}
H_{p_2} =& \frac{1}{\alpha}\frac{\partial}{\partial p_2}[N_1 + N_2 - D\frac{N_1}{1+N_1}] - \frac{\partial}{\partial p_2}\frac{N_2(w-c_2)}{D} \\
=& \frac{1}{\alpha}[\frac{\partial D}{\partial p_2} - \frac{\partial D}{\partial p_2}\frac{N_1}{1+N_1} - \frac{D}{(1+N_1)^2}\frac{\partial N_1}{\partial p_2}] - [\frac{\partial w}{\partial p_2}\frac{N_2}{D} + \frac{(w-c_2)}{D^2}(\frac{\partial N_2}{\partial p_2}D - \frac{\partial D}{\partial p_2}N_2)] \\
=& \frac{1}{\alpha}[-\alpha\frac{N_2}{1+N_1}\frac{1}{1+N_1} - \frac{D}{(1+N_1)^2}\alpha\frac{N_1N_2}{N_1+1}] - \frac{D}{1+N_1}\frac{N_2}{D} \\
& - \frac{(w-c_2)}{D^2}(-\alpha N_2 D + \alpha\frac{N_2}{N_1+1}N_2) \\
=& -\frac{N_2}{(1+N_1)^2} - \frac{DN_1N_2}{(1+N_1)^3} - \frac{N_2}{1+N_1} - \frac{\alpha(w-c_2)s_2}{D}(-D + \frac{N_2}{N_1+1}) \\
=& -\frac{N_2}{(1+N_1)^2} - \frac{DN_1N_2}{(1+N_1)^3} - \frac{N_2}{1+N_1} - \frac{N_2}{1+N_1}\frac{1}{D}(-D + \frac{N_2}{N_1+1}) \\
=& -\frac{N_2}{(1+N_1)^2} - \frac{DN_1N_2}{(1+N_1)^3} - \frac{N_2^2}{(1+N_1)^2 D} \\
<& 0
\end{aligned}$$

Appendix B

Appendix to Chapter 4

B.1 Evidence of learning r_{jt}

The demand estimation also gives product demand and cost states (d_{jt}, γ_{jt}). To see whether sellers indeed learn about r_{jt} from Market Insight, we first construct $r_{jt} = d_{jt} + (1-\alpha)\gamma_{jt}$, and then run the following diff-in-diff regression:

$$I_{jkt} = a_r \cdot r_{jt} \cdot MI_{p,kt} \cdot \text{after}_t + a \cdot X_{jkt}$$

where $MI_{p,kt}$ indicates whether seller k has Market Insight pro, after_t is 0 before the policy change, and 1 after. Covariates X_{jkt} contains $r_{jt}, MI_{p,kt}, \text{after}_t, r_{jt} \times MI_{p,kt}$ and seller characteristics $\tilde{X}_{kt}$. The parameter of interest is a_r: $a_r > 0$ means that Market Insight pro sellers start to carry higher profitability r_{jt} lipsticks after the policy change.

The results are shown in Table 9. Column 1 displays the estimates using linear probability model, controlling for seller fixed effect; column 2 displays the marginal effects from probit regression. First note that linear probability model and probit model give very similar estimates; the seller fixed effects don't affect estimates very much. Second, the treatment effect is positive and significant. It means that the policy change makes treated sellers carry higher profitability products. The treatment effect is also economically significant. If lipstick j's profitability r_{jt} is one standard deviation higher than mean, after the policy change Market Insight pro sellers' rate of carrying j goes up by 0.88%. Since the average rate of carrying any product is only 6.67%, this increase is equivalent to a 13% increase in the probability of carrying j.

Also note that $r_{jt}\times$MI pro has a negative coefficient. It means that Market Insight pro sellers, before policy change, react less to profitability than other sellers. This indicates that pro sellers have more noisy free signals, which is probably why they select into Market Insight pro in the first place.

Table B.1: Evidence of Learning r_{jt}

	Linear prob	Probit
r_{jt}	0.02411***	0.0238***
	(0.00017)	(0.00016)
r_{jt}×MI pro	-0.00896***	-0.0046***
	(0.00119)	(0.00129)
r_{jt}× MI pro × after	0.00730***	0.00782***
	(0.00172)	(0.00184)
MI pro		-0.03058***
		(0.00153)
after policy change	0.00174	0.00150
	(0.00190)	(0.00209)
control	Seller fe	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

- y : I_{jkt}, parameters on $\tilde{X}_{kt}$ not reported
- column 2 shows marginal effects
- in linear probability model, coef. on MI pro omitted because it's perfectly collinear with seller fixed effect; again we use non-switchers (of MI subscription) here.

B.2 Robustness checks

Longer panel

Using the whole sample from February to October 2019, the results in Table 8 are similar to that in Table 1: treated sellers' monthly revenue goes up by around 40%, and their rate of introducing new lipsticks goes up by around 7 percentage points compared to the nonsubscribers after the policy change. Here the placebo group, basic sellers sees lower monthly revenue. If the basic and pro sellers share similar time trends, then the treatment effect on pro sellers' monthly revenue is even higher than the 40% estimate.

Table B.2: Diff-in-Diff Effect, Longer Panel

	mon. revenue	rate of intro.
TREAT	0.42^{***}	0.07^{***}
	(0.09)	(0.02)
PLACE	-0.13^{**}	0.02
	(0.06)	(0.01)

- Use sample February through October 2019
- January ignored because "new" product undefined for the first month

Fat tails

In the online market, many sellers are very small. One might worry these tiny sellers drive the result. Here we exclude sellers who make less than 50,000 CNY ($\sim 7,500$ USD) revenue in January through October 2019.[1] The estimation results in Table 9 are still consistent with that in Table 1: treated seller revenue goes up by around 40% and their rate of introducing new lipsticks goes up by 7 percentage points.

Table B.3: Diff-in-Diff Effect, Drop Small Sellers

	mon. revenue	rate of intro.
TREAT	0.43^{***}	0.07^{***}
	(0.10)	(0.03)
PLACE	0.14^{*}	0.02
	(0.07)	(0.02)

- drop small sellers whose lipstick sales in 2019 $\leq 50,000$ CNY ($\sim 7,500$ USD)

Synthetic control

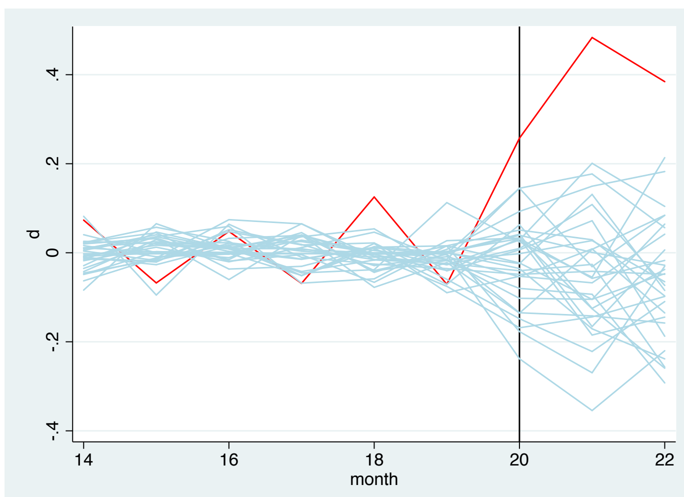
I divide the non-Market Insight sellers into 30 bins, according to their size measured in average monthly revenue. We also divide the Market Insight basic sellers into 2 bins according to their size. Then we assign weight to each of those 32 seller bins to construct the synthetic

[1]This is about the bottom 90th percentile of sellers; using other cutoffs give similar results. It is the common practice at Taobao economics team to drop the fat tail of tiny sellers. (In fact they suggest me to run this robustness check.)

control seller group to match the pre-policy change time trend. A key in synthetic control is long pre-treatment periods. Here we use the whole sample, from 2019/02 to 2019/10.[2]

First let's look at monthly revenue. The red line in Figure 3 is the treatment effect on pro sellers. While their monthly revenue is comparable with the synthetic control group before the policy change (2019/02 to 2019/07), it jumps up by nearly 40% right after the policy change. Note the magnitude is almost identical to that in previous specifications! The treatment effect is large not only economically, but also statistically. To show this, we calculate the treatment effect on each of the 32 not-treated seller groups: will the "treatment effect" defined on any other sellers be as big as on the treated sellers? The results are shown in blue lines (one blue line for each seller bin). It shows that none of the other groups has a treatment effect as big as the true treated sellers. In fact, treatment effects for non-treated sellers are smaller than half of that on pro sellers!

Figure B.1: Synthetic Control, Seller Monthly Rev.

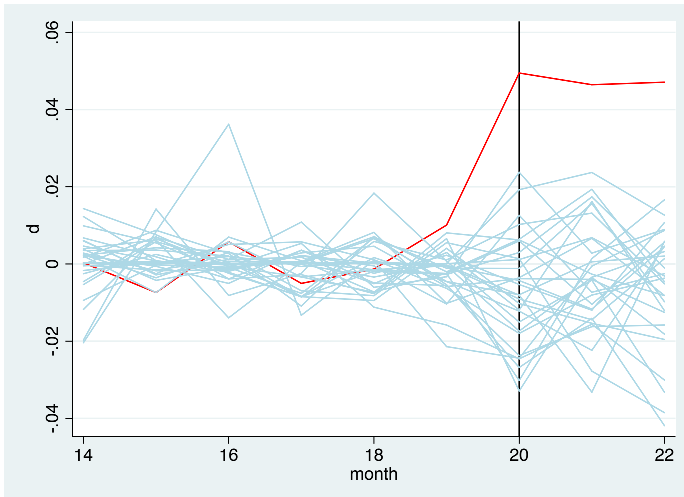


The same pattern holds for the rate of introducing new products. While the time trends are comparable for the pro sellers and the synthetic control group before the policy change,

[2] 2019/01 is dropped because the rate of introducing a new product isn't defined in this first month.

pro sellers raise this rate significantly immediately after the policy change. The treatment effect, around 5 percentage points, is again very similar in magnitude as that in previous analysis. This treatment effect is also quite significant statistically.

Figure B.2: Synthetic Control, Rate of Intro.



B.3 Gibbs sampling procedure

In this section, we describe the process of sampling from the joint posterior distribution, equations [20] and [21], using Gibbs sampling over $(d_{jt}, \gamma_{jt}, c_k, \alpha, \beta_d, log(q_{0t}), \beta_c, \Sigma_{uv})$.

G1. Sampling from the posterior distribution of $(\alpha^s, \beta_d^s, log(q_{0t})^s, \beta_c^s, \Sigma_{uv}^s)$

Given draws of $(d_{jt}, \gamma_{jt}, c_k)$ from iteration $s-1$, we begin iteration s by sampling $(\alpha^s, \beta_d^s, log(q_{0t})^s, \beta_c^s, \Sigma_{uv}^s)$. First, we sample β_c^s given $(\gamma_{jt}^{s-1}, c_k^{s-1})$. Given the linear form of the pricing equation and the multivariate normal prior, the posterior distribution of β_c is multivariate normal and the mean and variance can be expressed in closed form (Rossi et al., 2005, section 2.8) so it is simple to draw a value for β_c^s.

Then, we sample $(\alpha^s, \beta_d^s, log(q_{0t})^s)$ given β_c^s and d_{jt}^{s-1}. First, we deal with the zero-sales problem. When we observe sales quantity equal to zero, we draw $log(q)$ (LHS of equation [20]) from normal distribution $\mathcal{N}(log(q_{0t})^{s-1} + d_{jt}^{s-1} + \beta_d^{s-1} X_{kt} - \alpha^{s-1} log(p_{jkt}) + \mathbb{E}(u|v), \mathbb{V}(u|v))$ truncated below 0. The logic here is that sales ≤ 1 cannot be observed in reality; we should draw the sales according to the demand model. Note v in equation [21] can be constructed from the data given β_c^s and $(\gamma_{jt}^{s-1}, c_k^{s-1})$, given the s-1 iteration's variance-covariance matrix of (u, v), Σ_{uv}^{s-1}, we have $\mathbb{E}(u|v)$ as well as $\mathbb{V}(u|v)$. Second, we deal with the endogeneity of price in the demand equation. Conditioning on v effectively controls for the source of endogeneity in the demand equation (Rossi et al., 2005, section 7.1).

The final step in our use of the Gibbs sampler for $(\alpha^s, \beta_d^s, log(q_{0t})^s, \beta_c^s, \Sigma_{uv}^s)$ involves sampling Σ_{uv}^s given β_c^s and $(\alpha^s, \beta_d^s, log(q_{0t})^s)$. Again, the mean and variance of the posterior distribution have a closed form given the conjugate normal prior on the inverted Wishart prior on (u, v).

G2. Sampling from the posterior distribution of $(d_{jt}, \gamma_{jt}, \Sigma_j)$

Given $(\alpha^s, \beta_d^s, log(q_{0t})^s, \beta_c^s, \Sigma_{uv}^s)$ we next sample $(d_{jt}^s, \gamma_{jt}^s)$ and its variance-covariance matrix Σ_j^s for each lipstick brand in every month.

We first sample $(d_{jt}^s, \gamma_{jt}^s)$ conditional on $(\alpha^s, \beta_d^s, log(q_{0t})^s, \beta_c^s, \Sigma_{uv}^s)$, Σ_j^{s-1}, and c_k^{s-1}. This step uses the data and model parameters from the demand and pricing equations because (d_{jt}, γ_{jt}) enters into both these equations. At this stage we use Metropolis-Hastings accept/reject criteria product-by-product to sample from the posterior distribution.

Given $(d_{jt}^s, \gamma_{jt}^s)$ and $IW(I, 2)$ prior of $P(\Sigma_j)$, we could sample Σ_j^s.

G3. Sampling from the posterior distribution of (c_k, Σ_k)

Finally we next sample c_k^s and its variance Σ_k^s for each lipstick seller. We first sample c_k^s conditional on $(\alpha^s, \beta_d^s, log(q_{0t})^s, \beta_c^s, \Sigma_{uv}^s)$, $(d_{jt}^s, \gamma_{jt}^s)$, and Σ_k^{s-1}. This step uses the data and model parameters from the pricing equations; we use Metropolis-Hastings accept/reject criteria seller-by-seller to sample from the posterior distribution. Given c_k^s and $IW(I, 1)$ prior of $P(\Sigma_k)$, we could sample Σ_k^s.

Bibliography

[1] Ricardo Alonso and Odilon Camara. “Persuading Voters”. In: *American Economic Review* 106.11 (2016), pp. 3590–3605.

[2] Paul Asquith, Thomas Covert, and Parag Pathak. “The Effects of Mandatory Transparency in Financial Market Design: Evidence from the Corporate Bond Market”. Working Paper No. w19417. National Bureau of Economic Research. 2019.

[3] Jie Bai, Maggie Chen, and Daniel. Xu. “Search and Information Frictions on Global E-Commerce Platforms: Evidence from Ali-Express”. Working Papers 18-17, NET Institute. 2018.

[4] Patrick Bajari et al. “The Impact of Big Data on Firm Performance: An Empirical Investigation”. In: *AEA Papers and Proceedings* 109 (2019), pp. 33–37.

[5] Rajeev Batra and Indrajit Sinha. “Consumer-level factors moderating the success of private label brands”. In: *Journal of Retailing* 76.2 (2000), pp. 175–191.

[6] Dirk Bergemann and Stephen Edward Morris. “Information Design: A Unified Perspective”. Cowles Foundation Discussion Paper No. 2075R3. 2018.

[7] Jack (Xinlei) Chen et al. “An Empirical Investigation of Private Label Supply by National Label Producers”. In: *Marketing Science* 29.4 (2010), pp. 738–755.

[8] Pradeep K. Chintagunta, Andre Bonfrer, and Inseong Song. “Investigating the Effects of Store-Brand Introduction on Retailer Demand and Pricing Behavior”. In: *Management Science* 48.10 (2002), pp. 1242–1267.

[9] Allan Collard-Wexler, Gautam Gowrisankaran, and Robin S. Lee. “”Nash-in-Nash” Bargaining: A Microfoundation for Applied Work”. In: *Journal of Political Economy* 127.1 (Feb. 2019), pp. 163–195.

[10] Timothy G Conley and Christopher R Udry. “Learning about a new technology: Pineapple in Ghana”. In: *The American Economic Review* 100.1 (2010), pp. 35–69.

[11] Thomas R Covert. “Experiential and social learning in firms: the case of hydraulic fracturing in the Bakken Shale”. SSRN Working Paper No. 2481321. 2015.

[12] Gregory S. Crawford and Ali Yurukoglu. “The welfare effects of bundling in multichannel television markets”. In: *The American Economic Review* 102.2 (2012), pp. 643–685.

[13] Sanmay Das, Emir Kamenica, and Renee Mirka. “Reducing Congestion Through Information Design”. In: *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (2017), pp. 1279–1284.

[14] Michaela Draganska, Daniel Klapper, and Sofia B. Villas-Boas. "A Larger Slice or a Larger Pie? An Empirical Investigation of Bargaining Power in the Distribution Channe". In: *Marketing Science* 29.1 (2010), pp. 57–74.

[15] Anthony J. Dukes, Esther Gal-Or, and Kannan Srinivasan. "Channel Bargaining with Retailer Asymmetry". In: *Journal of Marketing Research* 43.1 (2006), pp. 84–97.

[16] Paul B. Ellickson, Pianpian Kong, and Mitchell J. Lovett. "Private Labels and Retailer Profitability: Bilateral Bargaining in the Grocery Channel". NBER Working Paper. 2017.

[17] Enric J. de Fortuny, David Martens, and Foster Provost. "Predictive Modeling with Big Data: Is Bigger Really Better?" In: *Big Data* 1.4 (2013), pp. 215–226.

[18] Andrew D Foster and Mark R Rosenzweig. "Learning by doing and learning from others: Human capital and technical change in agriculture". In: *Journal of political Economy* 103.6 (1995), pp. 1176–1209.

[19] Benjamin Golub and Evan Sadler. "Learning in social networks". SSRN Working Paper No. 2919146. 2017.

[20] Matthew Grennan. "Price discrimination and bargaining: Empirical evidence from medical devices". In: *The American Economic Review* 103.1 (2013), pp. 145–177.

[21] Zvi Griliches. "Hybrid corn: An exploration in the economics of technological change". In: *Econometrica* 25.4 (1957), pp. 501–522.

[22] Karsten Hansen, Vishal Singh, and Pradeep Chintagunta. "Understanding store-brand purchase behavior across categories". In: *Marketing Science* 25.1 (2006), pp. 75–90.

[23] S. J. Hoch, A. L. Montgomery, and Park Y. H. "Why the Private Label is one of the Few Brands to Show Consistent Long Term Growth". Paper presented at Allied Social Sciences Association Annual Meeting. Jan. 2002.

[24] Stephen J. Hoch and Shumeet Banerji. "When do private labels succeed?" In: *Sloan management review* 34.4 (1993), p. 57.

[25] Charles Hodgson. "Information Externalities, Free Riding, and Optimal Exploration in the UK Oil Industry". SIEPR Working Paper No. 19-013. 2018.

[26] Henrick Horn and Asher Wolinsky. "Bilateral monopolies and incentives for merger". In: *The RAND Journal of Economics* 19.3 (1988), pp. 408–419.

[27] Xiang Hui et al. "Certification, Reputation and Entry: An Empirical Analysis". Working Paper No. w24916. National Bureau of Economic Research. 2018.

[28] Nicolas Inostroza and Alessandro Pavan. "Persuasion in Global Games with Application to Stress Testing". Discussion paper, Northwestern University.

[29] Ganesh Iyer and J. Miguel Villas-Boas. "A bargaining theory of distribution channels". In: *Journal of Marketing Research* (2003), pp. 80–100.

[30] Ryan Kellogg. "Learning by drilling: Interfirm learning and relationship persistence in the Texas oilpatch". In: *The Quarterly Journal of Economics* 126.4 (2011), pp. 1961–2004.

[31] Anja Lambrecht and Catherine Tucker. "Can Big Data Protect a Firm From Competition". In: *Competition Policy International Antitrust Chronicle* (2017), pp. 1–8.

[32] N. Gregory Mankiw and Michael Whinston. "Free Entry and Social Inefficiency". In: *Rand Journal of Economics* 17 (1986), pp. 48–58.

[33] Sergio Meza and K. Sudhir. "Do private labels increase retailer bargaining power?" In: *Quantitative Marketing and Economics* 8.3 (2010), pp. 333–363.

[34] Fiona Scott Morton and Florian Zettelmeyer. "The strategic positioning of store brands in retailer-manufacturer negotiations". In: *Review of industrial organization* 24.2 (2004), pp. 161–194.

[35] Carlos Noton and Andres Elberg. "Are supermarkets squeezing small suppliers? Evidence from negotiated wholesale prices". In: *The Economic Journal* 128.610 (2018), pp. 1304–1330.

[36] Michael Ostrovsky and Michael Schwarz. "Information Disclosure and Unraveling in Matching Markets". In: *American Economic Journal: Microeconomics* 2.2 (2010), pp. 34–63.

[37] Koen Pauwels and Shuba Srinivasan. "Who benefits from store brand entry?" In: *Marketing Science* 23.3 (2004), pp. 364–390.

[38] Jagmohan S. Raju, Raj Sethuraman, and Sanjay K. Dhar. "The introduction and performance of store brands". In: *Management Science* 41.6 (1995), pp. 957–978.

[39] Jean-Charles Rochet and Jean Tirole. "Two-Sided Markets: A Progress Report". In: *The RAND Journal of Economics* 37.3 (2006), pp. 645–667.

[40] Gleb Romanyuk and Alex Smolin. "Cream Skimming and Information Design in Matching Markets". SSRN Working Paper No. 3178328. 2018.

[41] Ariel Rubinstein. "Perfect equilibrium in a bargaining model". In: *Econometrica* 50.1 (Jan. 1982), pp. 97–109.

[42] Marc Rysman. "The Economies of Two-Sided Markets". In: *Journal of Economics Perspectives* 23.3 (2009), pp. 125–143.

[43] Serdar Sayman, Stephen J. Hoch, and Jagmohan S. Raju. "Positioning of store brands". In: *Marketing Science* 21.4 (2002), pp. 378–397.

[44] Hank Schultz. *Private label sales hit 98 billion on 6 growth, report says.*

[45] Indrajit Sinha and Rajeev Batra. “The effect of consumer price consciousness on private label purchase”. In: *International Journal of Research in Marketing* 16.3 (1999), pp. 237–251.

[46] Andrew L. Steck. “Industry Dynamics with Social Learning: Evidence from Hydraulic Fracturing”. 2018.

[47] Robert L. Steiner. “The nature and benefits of national brand/private label competition”. In: *Review of Industrial Organization* 24.2 (2004), pp. 105–127.

[48] Oyvind Thomassen et al. “Multi-Category Competition and Market Power: A Model of Supermarket Pricing”. In: *American Economic Review* 107.8 (2017), pp. 2308–2351.

[49] Hal Varian. “Big Data: New Tricks for Econometrics”. In: *Journal of Economic Perspectives* 28.2 (2014), pp. 3–28.

Biography

Tiancheng Chen is a PhD candidate at Duke University and will complete his doctoral degree in economics in September 2021. Prior to the Ph.D., Tiancheng was an undergraduate student at the University of Hong Kong.