PRICE DISPERSION, MARKET COMPETITION AND REPUTATION:

EVIDENCE FROM CHINA'S ONLINE ELECTRONIC MARKET

by

Rentong Luan

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Spring 2017

PRICE DISPERSION, MARKET COMPETITION AND REPUTATION:

EVIDENCE FROM CHINA'S ONLINE ELECTRONIC MARKET

by

Rentong Luan

Approved: ____
James L. Butkiewicz, Ph.D.
Chair of the Department of Economics

Approved: ____
Bruce W. Weber, Ph.D.
Dean of the Alfred Lerner College of Business and Economics

Approved: ____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Michael A. Arnold, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

James G. Mulligan, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Joseph I. Daniel, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Xiang Qing, Ph.D.
Member of dissertation committee

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

I study price dispersion and the impact of market concentration and reputation using data collected from China's online markets for consumer electronics. The data provides not only the price information but also the recent sales volume for each seller. It shows that price dispersion does not diminish over time. Although I draw the same conclusion as the research on US market using the gap measurement between the two lists with lowest prices, I find the gap measurement itself is not an effective indicator. Using proper measurement of price dispersion, I find the dispersion is larger in larger markets, which contradicts the findings from data which do not include sales information. Moreover, I find the reputation and services provided by online sellers has little impact on their prices, except for the registered sellers.

Chapter 1

INTRODUCTION

The law of one price states that in an efficient market, identical goods have equal prices. Price dispersion, representing the variation across sellers and markets of the price of identical items, violates the law of one price. It has been an interesting topic at least since Jung (1960). He studied the prices quoted from Chicago automobile dealers for certain vehicles. The dealers located ten miles away from Chicago priced higher than the dealers within the city. Stigler (1961) found the investigation to be a great example for price dispersion of homogeneous goods. It is since, in many markets, consumers do not observe all the prices quoted at any given time. A buyer who wishes to determine the best price must search across sellers. The sellers take advantage of imperfect information and create price variation to extract surplus from consumers. Rothschild (1973) criticized Stigler's model in his survey paper arguing that fixed sample size search may not be optimal, and that the distribution of prices is not based on optimizing firm behavior. His survey on theoretical frameworks shows that different models analyzed have a variety of different equilibria; some are characterized by a single price, some by a distribution of prices.

Economists provided various search-cost models to explain the existence of price dispersion. Salop and Stiglitz (1977) set up a model with two different kinds of consumers, informed and uninformed. While the informed consumer always buy at the lowest cost, the uninformed consumers can only shop randomly. Assuming the stores have identical U-shaped cost functions, they show that a market equilibrium with price dispersion exists if there are enough uninformed consumers in the market. At the equilibrium, some stores will sell their products at a price equal to marginal cost and minimum average cost. Other firms produce lower output and sell at a higher price equal to the average cost at the lower output. The equilibrium holds only if there are enough uninformed consumer in the market to keep the stores selling at a higher price in business. In equilibrium, all firms earn zero profit and no firm will find it profitable to deviate from the pricing rule. Varian (1980) used the notion of temporal price dispersion. His model better represents the sales activities of retail stores. The stores vary their prices over time to stop the buyers from learning the experience. Nash equilibrium exists when all stores use the same random mixed pricing strategy. For each individual seller, promotional sales events are strategically employed over time to prevent consumer learning from experience. At each time point, a cross sectional view of the market shows dispersion. In his model, there are also informed and uninformed consumers. Sellers utilize sales events to discriminate against uninformed consumers in price.

Price dispersion is often examined with other market characteristics such as market competition and concentration. Stigler (1961) found that the dispersion of prices is negatively correlated with the stability of supply and demand. As the size of the market grows, agencies will be formed to collect and distribute information. Therefore, the cost of search will be smaller in a larger market. Reinganum (1979) proposes a model in which equilibrium price dispersion can be achieved for sequentially searching consumers. The model predicts that a decrease in search costs would lower the level of price dispersion as consumers' reservation prices fall. The high-cost firms will reduce their prices and the low-cost firms do not change their prices. Hence the price dispersion shrinks as search cost drops. On the other hand, MacMinn (1980) studies a model in which equilibrium price dispersion can be achieved for fixed sample searching consumers, provided search costs are sufficiently low. The price dispersion is supported and explained by the dispersion of product costs. The model predicts consumers would search more firms if the expected reduction in price is greater than the search cost. MacMinn (1980) shows that the variance of equilibrium price will eventually decrease with intensity of search.

In the last two decades, online shopping has become popular. Prices for the same item can be easily compared on shopping websites such as eBay.com and Amazon.com. Price dispersion is theoretically predicted to diminish as searching is less costly. The empirical results show the extent of price dispersion has become a popular research topic. Brown and Goolsbee (2002) found that increases in internet use significantly reduced the price and price dispersion of life insurance. There was no significant event effect for the startup of price comparison site nor for the new insurance types covered on the sites. On the other hand, Clay, Krishnan and Wolff (2001) found that online book prices are the same or lower than offline but the prices do not converge over time. Smith and Brynjolfsson (2001) also revealed a large amount of price dispersion for books across Internet retailers. They found the price dispersion may be caused by brand differentiation by sellers. They argued that consumers use brand information as a proxy for unobserved characteristics such as shipping reliability. Price dispersion persists over time whether shipping costs were considered or not.

The rise of e-commerce provides an opportunity for studying the determinants of dispersion empirically, such as competition and reputation. Yet the direction and magnitude of the impact were not consistent. Clay, Krishnan and Wolff (2001) studied the U.S. online book market and provided evidence that prices and price dispersion are lower for advertised or popular items. They found that dispersion is higher with more competitive firms. They also mentioned the potential differentiation from customer services but without data to support further investigation. Resnick and Zeckhauser (2002) empirically analyzed how eBay's reputation system works. The internet requires very little cost to provide and distribute customer feedback.

However, the incentive to provide feedback is small since feedback is a purely public good. And there is a concern of biased feedback because buyers tend not to leave negative feedback. The large percentage of positive feedback makes it hard to predict future behavior when there is no negative sign. Melnik and Alm (2002) studied the relationship between seller reputation and price using eBay.com online auction data. Their empirical finding is that the seller's reputation has a positive but small impact on the price. The effectiveness of eBay's feedback system is questioned.

Online shopping platforms, where consumers can compare the prices of seemingly identical products with a few clicks, have been studied recently. Lin and Chen (2014) investigated multiple online book websites in Taiwan, using a search-cost framework. They use advertisements and promotions as an indicator for lower search cost. Their empirical results show that prices and price dispersion are both lower for books that are advertised or popular. The observed normalized prices are lower when the number of big bookstores increases but higher when the number of fringe bookstores increases. Moreover, price dispersion significantly reduces when the number of competitors increases. Dispersion is smaller for the big bookstores when there are more big bookstores in a promotion period, if only big stores are considered. But the number of big stores has an insignificant impact for the fringe bookstores. Jolivet, Jullien and Postel-Vinay (2016) use a large transaction level dataset from one of the largest e-commerce platforms in France to estimate the effect of a seller's reputation on prices. The study focuses on books, CDs, video games and DVDs. The seller's reputation is included in the model. The results show a strong positive effect of seller reputation on prices. They also compare the reputation impact for professional sellers against the impact for regular sellers. The results show the reputation has a significant effect in both samples.

Among the empirical research of online markets, my research is most closely related to work by Baye, Morgan and Scholten (2004). My study follows their theoretical framework, and I investigate the interesting questions they covered as my starting point. Baye, Morgan and Scholten (2004) employed a "Spider program" to download price information from Shopper.com daily for a period of time. For each day, they tracked the listing prices of the top 100 popular products. They used an information clearing house model, which is suitable for mimicking a third party website that provides a list of prices charged by different firms in the market. Unlike search-cost models, the clearing house model assumes the costs for consumers to obtain price information are close to zero. Their main goal is to observe the change of dispersion in online markets over time. The measurement for dispersion they used is the percentage difference between the lowest two prices, the so called "price gap". They did not find meaningful results using other price dispersion measures. The empirical results of their research suggest that price dispersion in online markets is sizeable, pervasive, and persistent. Another goal of their research was to study the effect of market competitiveness on price dispersion. Their theoretical analysis shows that in a clearing house model price dispersion is greater in the small market than in the market with large number of competitors. Price competition increases as the number of firms in the market increases using the difference between the lowest two listing prices as the measurement of price dispersion.

My study contributes to this literature in several ways. While most empirical studies of internet markets focus on major US websites, little attention has been given to foreign E-commerce platforms. I use China's online market data to compare with the results from research on US markets. Taobao.com is one of the largest online marketplaces. Using a data abstracting program, I collected and cleaned over 5.8 million listing records for 93 items in 8 categories over a 3-month period. The amount of data is much larger than many studies on US websites and is comparable to the Jolivet, Jullien and Postel-Vinay (2016) study using data from PriceMinister.com. The comparisons highlight similarities and differences between US and Chinese online markets. Secondly, the most important benefit of using data from Taobao.com is that sales information is available. My empirical results show that price dispersion measure without sales information may be misleading. When sales information is included in calculations of price dispersion, the impact of the number of competing firms contradicts findings in literature on the US online market. The availability of sales information also enables me to study the market the impact of concentration on price dispersion. Moreover, I track the same products, instead of different ones, I am able to control for fixed effects of individual products. Baye, Morgan and Scholten (2004) collected the prices for the top 100 consumer electronic products and controlled for the fixed effects of ranking. However, without tracking the same products, they were unable to control for the fixed effects of individual product. Therefore, the findings in their paper are possibly due to the systematic difference between individual items rather than within each market. My empirical results show that controlling for fixed effects of individual products changes the conclusion of the impact from the number of competing firms.

The empirical data I collected has past sales information associated with every list, which enables me to test the robustness of Baye, Morgan and Scholten (2004). I find that price dispersion exists in all the online markets studied and that it does not diminish during the three-month period. In addition, the price gap measurement is negatively correlated with market size. These results are consistent with Baye, Morgan and Scholten (2004). However, I find that the price gap measurement itself may not represent the market characteristics well. The underlining assumption for price gap to be a good measurement for dispersion is that actual sales in the online market should be dominated by a few sellers who charge the lowest prices. My analyses show that the top sellers do not usually price significantly lower than their competitors. In fact, the top two sellers usually price near median of the prices listed in the market. The price distribution for different products can be quite different. Prices tend to be distributed uniformly in some markets, while they also can be highly concentrated in other markets.

Using sales weighted coefficient of variation as a measure of price dispersion, I find that price dispersion is actually larger with a larger number of competing firms. This is consistent with the results documented by Clay, Krishnan and Wolff (2001) but is different from the results of Baye, Morgan and Scholten (2004) and Lin and Chen (2014). The actual data show that the online markets are typically shared by a group of several sellers instead of just two sellers. The average market share for the top two sellers is 53.2% and the average market share for the top ten sellers is 80.6%. The market share of the top two sellers is negatively correlated with the number of firms. The extent of dispersion reduces for popular items, which is consistent with the evidence found by Lin and Chen (2014).

I find the impact of reputation of internet sellers is small in the Chinese market, which is consistent with Melnik and Alm (2002). It could be caused by the fact that Chinese online sellers usually have extremely high ratings so buyers cannot distinguish carefully. The services provided by the sellers also do not exhibit dramatic margins to the listing prices, except for sellers designated as registered or authorized business sellers. This designation has a positive and significant impact on price.

The rest of the dissertation is as follows. In Section 2, I review additional literature, including empirical results from other papers. In Section 3, I explain the background of Taobao.com and how the data were obtained and handled. In section 4, I analyze the actual data and compare my results against the findings on US online markets. Section 5 presents the conclusions and discussion of possible research.

Chapter 2

LITERATURE REVIEW

2.1 Price dispersion

Price dispersion, representing the variation across sellers and markets of the price of identical items, has been an interesting topic at least since Jung (1960). He studied the prices quoted from Chicago automobile dealers for certain vehicles. It was found that the prices quoted by telephone were competitive to the prices quoted in person. The dealers located ten miles away from Chicago priced higher than the dealers within the city. Stigler (1961) found the investigation to be a great example of price dispersion for a homogeneous goods. Prices change constantly unless all of the price quotes are acknowledgeable at a given time. The greater the instability of supply and demand conditions, the greater the dispersion of prices will be. Also, in any markets, there are always uninformed buyers at any time. The lack of information or experience creates dispersion. As the size of the market grows, agencies will be formed to collect and distribute information. Therefore, the cost of search will be smaller in a larger market.

The challenge to economic theory is how to describe a market equilibrium with price dispersion when at least some consumers are rational. Economists tried to explain the phenomenon disobeying the law of one price in different ways. One popular assumption is that consumers incur search costs as they shop. The sellers know that buyers need to pay for additional costs if they search for the lowest price. Thus, spatial price dispersion exists as the sellers have an incentive to price discriminate across buyers. Salop and Stiglitz (1977) set up a model with two different kinds of consumers, informed and uninformed. While the informed consumer can always buy at the lowest price, the uninformed consumer can only shop randomly. Assuming the stores have identical U-shaped cost functions, a market equilibrium with price dispersion exists if some assumptions over parameters hold. At the equilibrium, some stores will sell their products at a price equal to marginal cost and minimum average cost. Other firms produce lower output and sell at a higher price equal to the average cost at the lower output. The equilibrium holds only if there are enough uninformed consumer in the market to keep the stores selling at a higher price in business. Once the equilibrium is reached, no firm will find it profitable to deviate from the pricing rule. However, in their model, there are some stores that always charge a lower price than the others. Therefore, if consumers can learn from previous shopping experience, the equilibrium will eventually collapse to the competitive equilibrium. Wilde and Schwartz (1979) also used the spatial price dispersion in their model and predicted that low search costs will make consumers better off. They start with the consumer's strategy of shopping. They argue that sequential searching may not be optimal because consumers can always stop when they find price variation is not large, given they do not have the full knowledge of the distribution of prices in market. A better searching strategy will be combining sequential searching and the fixed sample search, because some consumers enjoy visiting stores and always choose to visit at least two stores before buying. The key parameter of their model is the ratio of searching shoppers to the total number of consumers in the market. If the ratio is small, the equilibrium price distribution is continuous on some interval. If the ratio is relatively large, a mass point emerges at the competitive price.

Alternatively, Varian (1980) used the notion of temporal price dispersion. That is, the stores vary their price over time to stop the buyers from learning from experience. His model better predicts the sales activities of retail stores. Nash equilibrium also exists when all stores use a mixed pricing strategy. At each time point, a cross sectional view of the market shows price dispersion. But for each individual seller, over time promotional sales events are strategically employed to prevent consumer from learning from experience. In his model, there are informed and uninformed consumers and stores use randomized pricing strategies. Sellers utilize sales event to price discriminate between informed and uninformed consumers. Although it is not included in his model, he mentioned a more realistic case should consider inventory costs, cyclical fluctuations, loss leader behavior, advertising behavior and so on.

However, it is commonly agreed and intuitive tool to model newspaper advertising behavior. One tweak of his model is to incorporate the information cost as endogenous. Under this assumption, consumers can pay a fixed cost to be informed of all prices. If there is some search cost difference between certain groups of people, price dispersion persists in the market.

Reinganum (1979) proposes a model in which an equilibrium price dispersion can be achieved with sequentially searching consumers. The key assumptions are consumer demand is downward sloping and firms have heterogeneous marginal costs. In the model, consumers sequentially search and update the expected price distribution. Each search costs a fixed amount. Consumers must weigh the expected benefits against search cost. The model predicts that a decrease in search costs would lower the level of price dispersion as consumers' reservation prices fall. The high-cost firm will reduce their prices and the low-cost firms do not change their prices. Price dispersion shrinks.

MacMinn (1980) studies a model in which equilibrium price dispersion can be achieved for fixed sample searching consumers, provided search costs are sufficiently low. The price dispersion is supported and explained by the dispersion of product costs. The model predicts consumers would search more firms if the expected reduction in searched price is greater than the search cost. He shows that the variance of equilibrium price will eventually decrease with intensity of search.

The Internet challenges theoretical economic models which use search cost to explain price dispersion. In online markets the search cost to compare prices is close to zero or at least much lower than before. New generations are familiar with the Internet and online shopping. Economists have used econometric techniques in addition to theoretical models to analyze price dispersion. The technology growth provides both questions and new tools to answer these questions. Although the traditional economic view suggests the increasing use of the Internet would increase market competition, there is not much empirical research to support the argument. One positive example of a drop in price dispersion is from Brown and Goolsbee (2002). They found that the increase in internet use significantly reduced the price and price dispersion for life insurance. They studied how internet comparison shopping sites affected life insurance in the 1990s. In their study, individual insurance policies and policy characteristics are known. Given that, the increases in internet use significantly reduced the prices of term life insurance, by 8-15%. Moreover, they found there was no significant drop before the internet comparison site began or before insurance was sold online. A key finding is they firstly compared the impact of internet competition on prices and price dispersion in traditional off-line markets. Life insurance is a high search cost and relatively high markup product. In the mid-90s, a group of price comparison sites began and it shows significant impact in reducing life insurance price dispersion. On these sites, customers would answer a medical questionnaire online including the desired coverage and get quotes from different companies at the same time. The website did not sell life insurance directly. Since additional connections, such as blood test, are still taken offline, the websites are nearly strictly used as a search engine. They adopted the sequential searching model with internet users as informed consumers who pay zero or very little search cost. The average price drops monotonically as the proportion of informed consumers increases. However, the price dispersion is not linearly related to search costs. If all consumers are informed, then all sellers charge the same competitive price. If none of the consumers are informed, then all sellers charge the monopoly price. Therefore, when the proportion of informed consumer increases from zero to one, price dispersion should first increase and then eventually fall.

Others have found that price dispersion is persisting as internet use increases. Smith and Brynjolfsson (2001) found evidence of substantial price dispersion for books across Internet retailers. They stated that the persistent price dispersion may result from brand differentiation across online retailers. In their sample, the top three brands can charge $1.72 higher than generic sellers for a homogenous product. They argue that consumers use retailer brand information as a proxy for unobserved characteristic such as shipping reliability. The data they used are panel data of price search sites in the market for books. What is noticeable is that they link consumer information, consumer cookies and the sorting field, with the click-through information. In the setup of shopping procedure, consumers first choose which homogenous product to buy, then choose from listed sellers given their shipping time, shipping services and so on. Consumers can sort the list by desired field. However, whether the consumer brought the book or not is unknown. They used last webpage visited as the consumer's final choice. Price dispersion is observed to persist. And majority of consumers were not choosing the lowest price as their final choice. An empirical nested multinomial logit regression was used where consumers first choose between big brands and generic retailers. Then they sequentially choose a specific retailer and finally choose shipping options. Results show that consumers who use search engines to shop respond to well-known retailers. In particular, consumers who care more about shipping time are likely to select from well-known retailers.

Others also have studied firm pricing strategies but do not have information on sales or on consumer behavior. Clay, Krishnan and Wolff (2001) found that online book prices are the same or lower than offline but the prices do not converge over time. They collected data over time for 32 online bookstores. Each bookstore has an independent website. Among the big three brands, Amazon was 5 percent more expensive than Barnes & Noble and 11 percent more expensive than Borders. During the period of the study, prices for different types of books were substantially above cost and did not drop. Although the standard deviation decreased, price dispersion persisted. To explain the price dispersion, they focused on two reasons. The first is that consumers do not have perfect information about prices. Second is that the product is not identical because it bundles the item itself as well as services. Bookstores try to differentiate themselves from others in numbers of ways. This leads to the discussion of firm strategy. By examining the normalized average prices and standard deviations using fixed effects models, they found that small sellers tend to price their product following dominating sellers. Some small stores specialize to avoid competing with big sellers. For big sellers, online stores were considered as an advertisement tool to supplement physical stores. They also notice there is differentiation in terms of customer services but without supporting data.

One paper is key to my study. Baye, Morgan and Scholten (2004) examined daily prices on a price comparison site, Shopper.com. They employed an information clearing house model where a third party provides a subset of consumers with a list of prices charged by different firms in the market. The empirical evidence suggests that price dispersion in online electronic markets are sizeable, pervasive, and persistent. The price dispersion increases when there are more lists for the product. Comparing my result with the results from Baye, Morgan and Scholten (2004) will show the similarity and difference of online markets in China and the US.

2.2 Market competition and concentration

The relationship between price dispersion and market concentration has been widely investigated. The general finding in literature is that high concentration is associated with significantly higher prices. Singh and Zhu (2008) tested how the prices change with the number of competitors in the market. They found that the number of competitors have significant impact over prices. They also indicated endogeneity could be a severe problem in price-concentration regressions. They introduced a two-stage estimation procedure in which an equilibrium model of endogenous market structure provides correction terms for the second-stage price regression. The endogeneity problem was also noticed by Evans, Froeb and Werden (1993). They pointed out two possible reasons. First, performance feeds back into structure, causing a simultaneous equations bias. Second, the measured concentration, outputs or revenue, are correlated with error terms in the price regression. Use of panel data and instrumental variables can solve both issues. Their empirical finding showed the bias is substantial and negative.

Some other literature focuses on internet auctions. Bajari and Hortaçsu (2003) examined an internet auction dataset of coin auctions to explore the determinants of bidder and seller behavior. They specify and estimate a structural econometric model of bidding on eBay. They measure the effect of entry cost associated with bidding and simulate profit-maximizing seller revenue under different reserve prices. The empirical data show that bidders usually submit their bids close to the end of the auction and sellers tend to set minimum bids at levels considerably less than the items' book values.

Borenstein and Rose (1991) studied price dispersion that US airline companies charging different customers. They found the expected variation between two passengers can be 36 percent of the average ticket prices. Furthermore, the dispersion is larger on the routes with more competition or with lower flight density, controlling for variations from cost impact. Additional information used to illustrate the true dispersion include population, product differentiation and market characteristics. The expected effect of market structure on price dispersion follows classical theory. That is, price dispersion increases with concentration if the market is close to oligopoly and decreases with concentration if the market is close to monopolistic competition. Empirical results show that a carrier's price dispersion within a market increases when the number of competitors increases. They also found that a tourist route has less price dispersion compared to a business route. The reason is that airlines tend to attract the high-fare business travelers with loyalty plans for greater long-term revenue. This finding suggests that for products that consumers tend to buy more frequently, some sellers try to set their prices away from the average price to attract customers for repeat purchases. While if the product is more durable, sellers usually do not price lower since it is hard to tell whether the discount will be offset by future sales.

Lewis (2008) directly modeled dispersion as a function of the density of local competition and other seller or market characteristics. He measured price dispersion among gas stations, while these stations were considered as differentiated sellers with unobserved fixed effects. Significant price dispersion exists even controlling for differences in station characteristics, and price differences between sellers change frequently. In his research, a fixed effects model was adopted to control for unobserved seller heterogeneity. Sellers are expected to change their relative prices to prevent consumers from learning the equilibrium price distribution when the product is repeatedly purchased by the same consumer. The extent of price dispersion is related to the extent of competitiveness, but this relationship varies significantly depending on the type of seller and the composition of its competitors. The estimate between seller density and dispersion is strongly negative among small sellers, but is insignificant or weakly positive among big sellers. A key result, which is different from other research, is that price dispersion is measured relative to nearby sellers rather than the entire city. He believes that price dispersions are caused by difference of seller characteristics and consumer heterogeneity. To be more specific, buyers that are more willing to purchase from low-brand sellers are also willing to seek the lowest price possible. While buyers who are not willing to purchase from low-brand sellers are also less likely to search for the lowest price. Therefore, both low-brand sellers and high brand sellers will survive in the market and have some consumers to purchase from them.

Barron, Taylor and Umbeck (2004) also find that price dispersion among gas stations is negatively correlated with the number of sellers in a local market. They concluded that price dispersion is consistent with models of spatial competition rather than models of imperfect information and consumer search. In their spatial monopolistic competition model, two reasons can drive price dispersion in equilibrium. The first is seller's heterogeneous demand, which means that the visiting cost could be different for consumers even each seller's marginal cost is the same. The second one is sellers' marginal production costs could be different. Under either condition, a local market for gas stations can reach equilibrium by adding certain reasonable assumptions. For potential issues with the research, they did mention the potential endogeneity of seller density. However, Barron, Taylor and Umbeck (2004) used observed station characteristics to control for seller heterogeneity, leaving the possibility that unobserved station differences are responsible for some of the remaining price dispersion.

In more recent research, Lin and Chen (2014) investigated multiple online book websites in Taiwan, using a search-cost framework. They use advertisements and promotions as an indicator for lower search cost. Their empirical results show that prices and price dispersion are both lower for books that are advertised or popular. The observed normalized prices are lower when the number of big bookstores increases but higher when the number of fringe bookstores increases. Moreover, price dispersion significantly reduces when the number of competitors increases. Dispersion is smaller for the big bookstores when there are more big bookstores in a promotion period, if only big stores are considered. But the number of big stores has an insignificant impact on the fringe bookstores sample. Their findings on the impact of market competition on dispersion is different from the results documented in Clay, Krishnan and Wolff (2001).

In recent research, An, Baye, Hu, Shum and Morgan (2015) used UK data for PDA to present a general model of online price competition. Their results suggested that competitive effects in this online market are more closely aligned with the simple homogeneous product Bertrand model than might be expected given the observed price dispersion and number of firms. If two firms remain in the market post-merger, the average transaction price is roughly unaffected by horizontal mergers. However, there are potential distributional effects among price sensitive shoppers and loyal shoppers. Notice the number of competing firms is unknown or in dispute here.

2.3 Reputation

One important factor for online price is the seller's reputation. Unlike the offline stores, online store quality is not observable to buyers thus historical reputation usually is the only source for buyers to judge sellers. Most popular online platforms use buyer feedback to measure reputation. Theoretical models predict a positive relationship between a seller's reputation and buyers' willingness to pay. Some empirical studies support the argument. Houser and Wooders (2006) found the seller's reputation is both economically and statistically significant. In their model of eBay auctions, there is a single seller and *n* bidders. Both seller's and buyers' reputations are publicly known. The bidding price is privately known. The seller must decide whether to default on the auction contract, evaluating the risk of the winning bidder defaulting. Under certain conditions, second highest price can be regressed on seller's reputation, auction and product characteristics. They also pointed out that both sellers and buyers tend to behave well for future gain. However, whether the reputation system is a good predictor for future performance is unconfirmed in the research.

There are more empirical papers on seller reputation in eBay auctions. Lucking-Reiley, Bryan, Prasad and Reeves (2007) studied coin auctions on eBay and found that negative feedback has a strongly negative effect, while positive feedback has a small positive effect. It is noticeable that they found that the eBay summarized seller's rating score has no significant impact on the price. Buyers are sensitive to the number of negative feedback rather than a system generated, ambiguous rating. Minimum bids, reserve prices and auction periods have positive effects on the final auction price. However, the effectiveness of a seller's strategy is hard to measure since aggressive strategies may prevent the good from being sold at all. A more effective strategy would be setting the end of auction time to weekend, when there are more potential buyers surfing online.

Melnik and Alm (2002) discussed why the seller's reputation is important in determining buyer bids. Their empirical finding is that the seller's reputation has a positive but small impact on the price. They mentioned potential defect of using eBay system calculated ratings due to several reasons. First, not every single transaction generates a feedback. Second, buyers have little incentive to leave feedback, especially for those who are somehow satisfied. Also, sellers can always change identity. There is no real standard way to distinguish a seller's fraudulent action from honest mistakes. Even nowadays eBay and other websites start to restrict a seller's ability to open multiple stores, it is hard to eliminate that when small business usually have more than one participant. Another interesting topic briefly mentioned in this paper is that eBay went to court to prevent other auction sites from using feedback information on eBay as an advertisement.

Resnick and Zeckhauser (2002) empirically analyzed how eBay's reputation system works. The target is to understand why the reputation system works. For online retailers, the cost of providing and distributing feedback information is relatively low. However, the incentive for customers to provide feedback is also small since feedback is a purely public good. Nearly half of the buyers do not provide feedback. Also, there is a concern of biased feedback because buyers tend not to leave negative feedback. A surprisingly large fraction of feedback is positive for most sellers. Resnick and Zeckhauser characterize this phenomenon as a high courtesy equilibrium, in which people would like to do the right thing as little cost is associated. They have several interesting findings. Current reputation profiles were predictive of future performance. However, the large percentage of positive feedback makes it hard to predict future behavior. And they do not find significant price premiums generated by a better reputation. It only helps the sellers to sell more. The high correlation between seller and buyer feedback indicates that players in the "trust between strangers' game" is interactive.

Related research, not using eBay data, focuses on the relationship between price premium and reputation. Landon and Smith (1998) studied the impact on price of product quality and reputation using data for Bordeaux wine. A general result shows that the price premium far exceeds the improvement of expected quality. Both individual firm reputation and market reputation are important. But impact on price is disaggregated into individual firm reputation, while market reputation is only valued as a predictor of future quality. A key difference of this paper is it takes both collective reputation and individual firm reputation into consideration. This is suitable when acquiring accurate and comparable information of wine providers is relative difficult. In other words, the nature of the product quality depends on judgment and is costly to evaluate. Thus, the collective reputation, which is measured by the average quality of a specific group of providers, is relative easy to acquire. Empirical evidence also indicates that consumers consider a long-term reputation for quality to be a better indicator than recent reputation movements.

In most recent research, Jolivet, Jullien and Postel-Vinay (2016) use a large transaction level dataset from one of the largest e-commerce platforms in France to estimate the effect of a seller's reputation on prices. The study focuses on books, CDs, video games and DVDs. The results show a strong positive effect of seller reputation on prices. They also compare the reputation impact for professional sellers against the impact for regular sellers. The results show the reputation has significant effect in both samples.

Chapter 3

BACKGROUND AND DATA

3.1 China’s E-Commerce Platform

The data for my research were collected from the dominating Chinese online commercial website, Taobao.com. Established in 2003, Taobao.com took just one year to rank as the top shopping platform in the Chinese online market. Its market share was about 80% of Chinese e-commerce in 2010. Taobao.com held 46.9% of the B2C (Business to Consumer) market and 90.5% of the C2C (Consumer to Consumer) market at that time. As the transaction size reached 208 billion RMB and the number of registered buyers exceeded 170 million in 2009, Taobao.com became one of the largest marketplaces in the world. Taobao.com sales are now equal to those of Amazon.com and eBay.com combined. It is completely free for regular sellers, except for a relatively small amount of deposit. It owns a third-party payment system, like Paypal.com. Buyers can also pay by credit cards for most B2C items and some C2C items. From many aspects, Taobao.com is the ideal source to study the online market in China.

There are abundant data on Taobao.com to be collected. By the end of 2008, there were 176 thousand sellers and over 8 million items listed. These lists are arranged into 23 categories and 117 sub-categories. Starting from the homepage, one can reach the desired item by searching the name of the item or by browsing a specific category. If the keywords entered are not clear enough, the webpage will automatically suggest more keywords. As soon as the description of that item is sufficiently clear, the buyer will see a page with numbers of lists posted by different sellers. These lists can be ordered by price, number sold recently, seller's reputation or number of times being browsed. Each listing includes the seller's name and location, price, shipping cost, number sold in last thirty days, number of comments and additional services provided. Buyers can browse these briefs and decide which one to click on. Then buyers will be linked to a page for that list with more information.

Figure 3.1 is the snapshot for an item search result webpage.

Like the other electronic platforms, paying an advertisement fee allows sellers to be listed at the top of the search results. Given the mechanism of Taobao.com, each of the top sellers is placed on top of buyers' search with some probability. If the seller would like to pay for advertisement, the chance of being place on top become larger. But it is not guaranteed to be placed on top of the search every time.

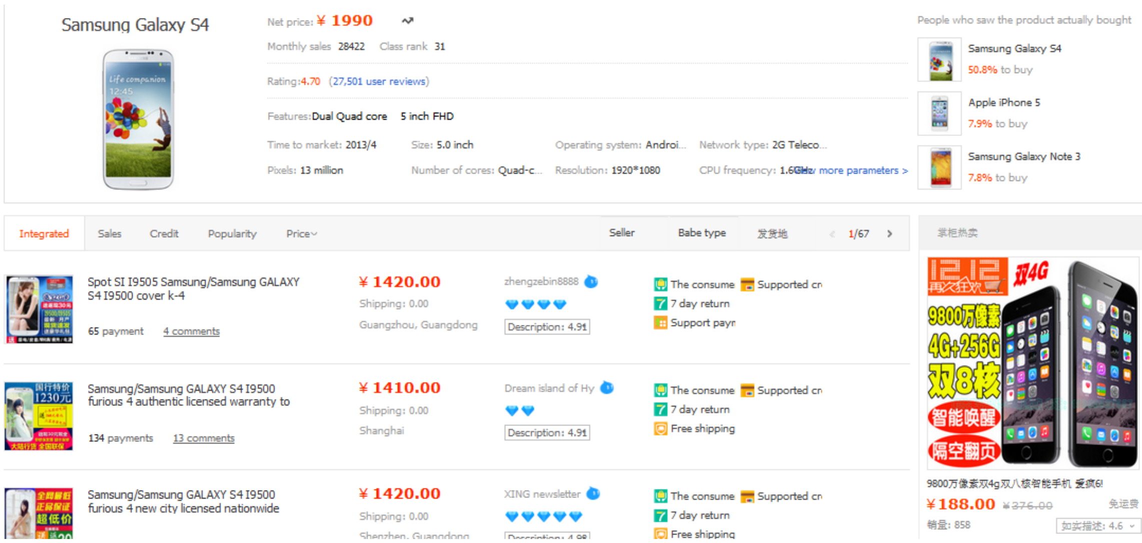


Figure 3.1 Snapshot of item search result page

Within each list, additional information is provided. Different buying options for the item are shown at the center of the page, including package and color. The specification of the item will be displayed. The seller's detailed reputation, including number of positive and negative feedback, consistency of description, service attitude and quickness of shipping will be shown on top. Buyers can also browse the recent comments and recent transaction records in this page.

Figure 3. is the snapshot for a list.

To my knowledge, Taobao.com has multiple rules to prevent cheating activities by sellers. First, multiple lists are not allowed and will be severely penalized. Taobao.com checks for multiple lists at the time a new list is being created. Second, each payment account is linked with a bank, which is linked with the unique personal

ID number. Thus, the possibility of false transactions, such as fake trading with friends and family to build up reputation, is greatly lowered.

Figure 3.2 Snapshot of item list page

3.2 Data collection

Using methods like those used by Baye, Morgan and Scholten (2004), the "Spider Program", I download the data from the website automatically for three months. It does not only have price data but also the quantity sold. I used a free website data mining system called "Locoy Platform" to obtain the data. I captured all the visible text and automatically stored it in Microsoft Access. The Locoy system reads the website in PHP language and save the useful information. Then I used the embedded searching tools to save the useful information into a dataset. I notice this method is widely used in business analysis services for sellers to adjust their pricing strategies.

The data are retrieved from June 16th, 2013 to September 25th, 2013. The program ran automatically after midnight and ended before noon. I checked the outputs every day to ensure the programs generated expected results. One difficulty I faced was that the layout of the website was updated irregularly. I spent some time to reprogram to continue retrieving data. In the end, the data between July 25th, 2013 and August 11th, 2013 as well as between September 11th, 2013 ad September 13th, 2013 were unusable when I reprogrammed the extraction code. That left me with usable data from 77 days between June 16th and September 25th.

Unlike Baye, Morgan and Scholten (2004), I used a fixed list of items instead of tracking most popular items every day. Tacking the most popular items can give large amount of sales but not the most similarities. Popular items can change frequently. Items on the top of popular lists only show up at the peak of the product lifetime. Instead, tracking the same items preserves information about how the popularity changes impact on competition and pricing of the item. I chose 8 categories and the top 15 items for each category. Items are ranked by quantity being sold in last month.

Because of data quality issues, I remove twenty-seven items from my study, leaving a total number of 93 products. The number of lists for a particular item is limited by Taobao.com to 4,000. This could result in an incomplete information of all listed prices. However, none of the items has more than 4,000 active lists. In other word, no item has more than 4,000 sellers with non-zero number of sales in last 30 days. The pricing information is sufficient for analysis at least for active markets. I also recorded seller reputation information if it is available.

Figure is the snapshot of a seller's store webpage.

Figure 3.3 Snapshot of store reputation page

Historical reputation information is not available for all sellers. For those having reputation information, registered sellers have so called dynamic reputation information, such as relative speed of shipping, displayed on their store web pages. The regular sellers, on the other hand, have their total historical feedback and number of positive feedback shown on their store web pages. I merged reputation information into the listing information to get total of 6,313,479 raw observations. Each observation contains the information for a given list in a day.

Table 3.1 Example of data structure

Date	Category	Name	TrueRank	WeekAvgPrice	WeekTotSale	NumList	Price	Shipping	30DaySale	SellerID	ItemID	Location	Tmall	DescripRate	AttitudeRate	SpeedRate	DescripRelative	AttitudeRelative	SpeedRelative	DescripRelativeRate	AttitudeRelativeRate	SpeedRelativeRate
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	57	6	24	7E+08	####	Nanchang, Jiangxi	1	4.8634	4.86108	4.871				0	0	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	65	5	0	2E+07	####	Shenyang, Liaoning	0	4.95507	4.96025	4.953	over	over	over	75.72	72.84	69.59
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	60	30	2	7E+08	####	Shanghai	1	4.82898	4.87026	4.896				0	0	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	62	0	6	7E+08	####	Changsha, Hunan	1	4.82903	4.83185	4.853		fair		0	0	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	67.9	0	55	7E+08	####	Shenyang, Liaoning	1	4.83107	4.83046	4.823		fair	fair	0	0	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	59	0	390	7E+08	####	Shenzhen, Guangdong Province	1	4.83456	4.84501	4.839				0	0	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	60	10	1	7E+08	####	Beijing	1	4.8703	4.86166	4.889				0	0	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	62	10	0	7E+08	####	Changsha, Hunan	1	4.8139	4.81776	4.835		fair		0	0	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	69	0.01	0	7E+08	####	Fuzhou, Fujian	1	4.76634	4.78085	4.827	lower	lower	fair	0.74	1.09	0
20130617	Flash Drive	Kingston DT101G2(16GB)	1	58	6982	4819	58	10	12	4E+08	####	Shanghai	1	4.83024	4.83651	4.826			fair	0	0	0

Table 3.1 is a sample of how data was stored. Name is the physical name of the item. True rank is the current rank of the item in its category. Weekly average price, weekly total sales and number of lists are provided by Taobao.com at each item search result webpage. These can be used to identify the popularity of the item. However, the weekly average price does not contain the shipping cost information. Prices and shipping costs are in RMB (￥). Sales are the number of items sold during the last 30 days. Tmall is the indicator of whether the seller is a registered seller or not. Description rating, attitude rating and speed of shipping rating are 30-day dynamic rating with values between 0 and 5. Description rating can be interpreted as the accuracy of the item description provided by the seller. Sellers can be punished by the low rating if they try to over-advertise the item. Attitude rating is rating of customer service. Speed of shipping rating measures the satisfaction rate of delivery time. Description relative rating, attitude relative rating and speed of shipping relative rating compare the three ratings above with average ratings for sellers in the same category. Each of these relative ratings contains the sign of comparison, which has the value of “lower’, “fair” and “over”, and the relative percentage. For example, description relative rating “over 75.72” means this seller’s accuracy of description is higher than 75.72% competitors.

Taobao.com does monitor irregular prices and removes lists with unreasonable prices. However, I still observe some irregular pricing lists. Based on the research from sellers' online forums, there are two major reasons for over pricing. First, a new seller needs at least 10 lists to open a new store. Considering a new seller who has the resources for only one item, if he lists another nine items that are not available, it is better to list an extremely high price so no one will buy it. Second, some small sellers also adjust their prices when their inventories are low. The reason is that if the seller removes the list, the search ranking of the list will be dropped when it is re-listed. In the end, some sellers use price adjustment to temporarily cease their lists.

I use 25% and 400% of weekly average price as the cut off for outliers. As a result, 6,313,479 observations were reduced to 6,199,832 observations. Then I examine the data by checking them in Excel and remove some problematic records. For example, I eliminated the dates with the number of lists much more than those of the two neighboring dates. Such records are considered as defect and should not be included in the research. In the end, 5,831,215 number of records were finally selected for further analysis.

Chapter 4

EMPIRICAL RESULTS

4.1 Price distribution

Online platforms require little cost to quote prices, thus prices are expected to converge. However, does this actually occur? Before moving to study dispersion, I review the price distributions for different markets. I pick one item from each category as an example. Below are their price distributions on Jun 16, 2013. The price distributions vary across different items. They can be roughly categorized into two groups.

4.1.1 Case study on highly concentrated distribution

The first group contains headphones and speakers. Interestingly, the two items are made by the same manufacturer. In this group, prices are extremely concentrated in a fairly narrow price band.

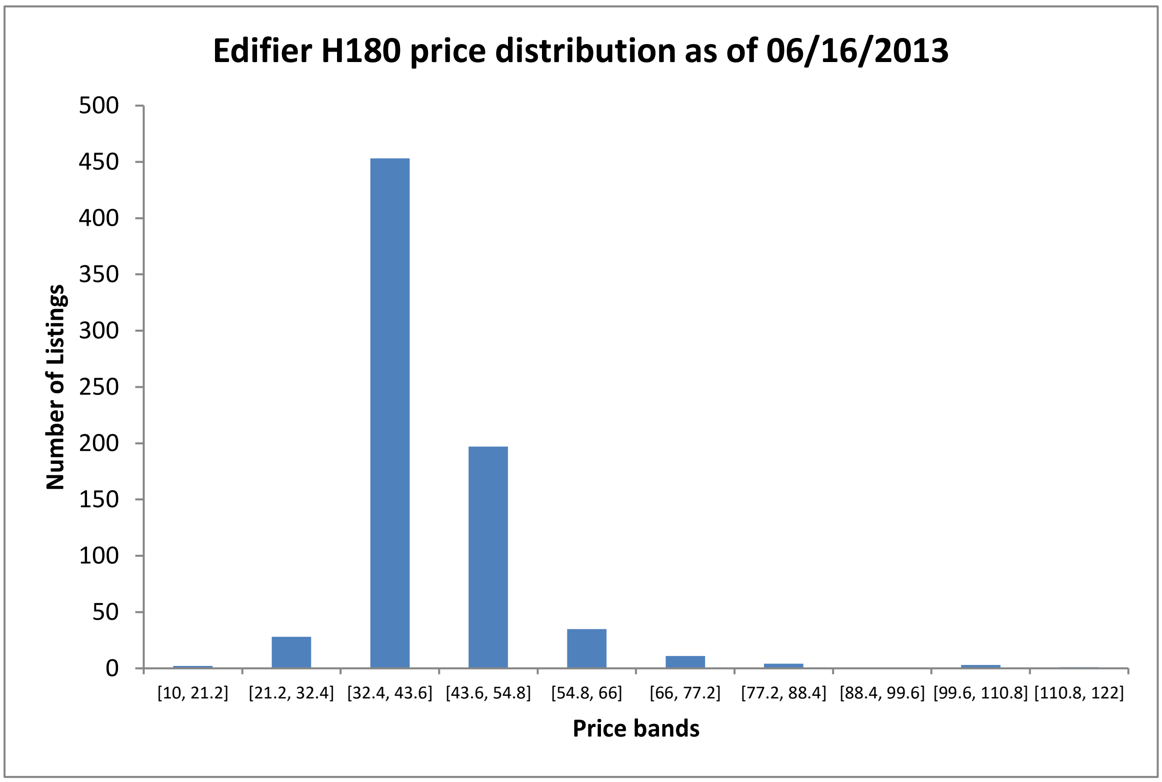


Figure 4.1 Edifier H180 price distribution as of 06/16/2013

As we can see from the chart above, though the listing price ranges from ￥10 to ￥122 for the same item, most list prices are concentrated in [32.4, 54.8] band.

There are 734 sellers in the market. 453 sellers charge prices between ￥32.4 and ￥43.6, which are 61.7% of the total number of sellers. Another 26.8% of all the sellers charge prices between ￥43.6 and ￥54.8. These two groups of sellers are together 88.6% of the total number of sellers of the market. It is also where the top ranked seller listed its price. The low sales price leaves little room for sellers to compete with the top ranked seller on price.

Notice the above chart includes all prices listed on the market. It is interesting to check the price distribution with sales. If the market is actually closer to perfect competition rather than monopoly, then removing the listings with no sales should make the distribution even more concentrated and reduce the average price. The average price in Figure 4.1 is ￥42.9, with standard deviation of 9.34. In Figure 4.2, I removed the lists with zero or only one sale in the last thirty days. The reason to remove the lists with only one sale is just to reduce potential data noise.

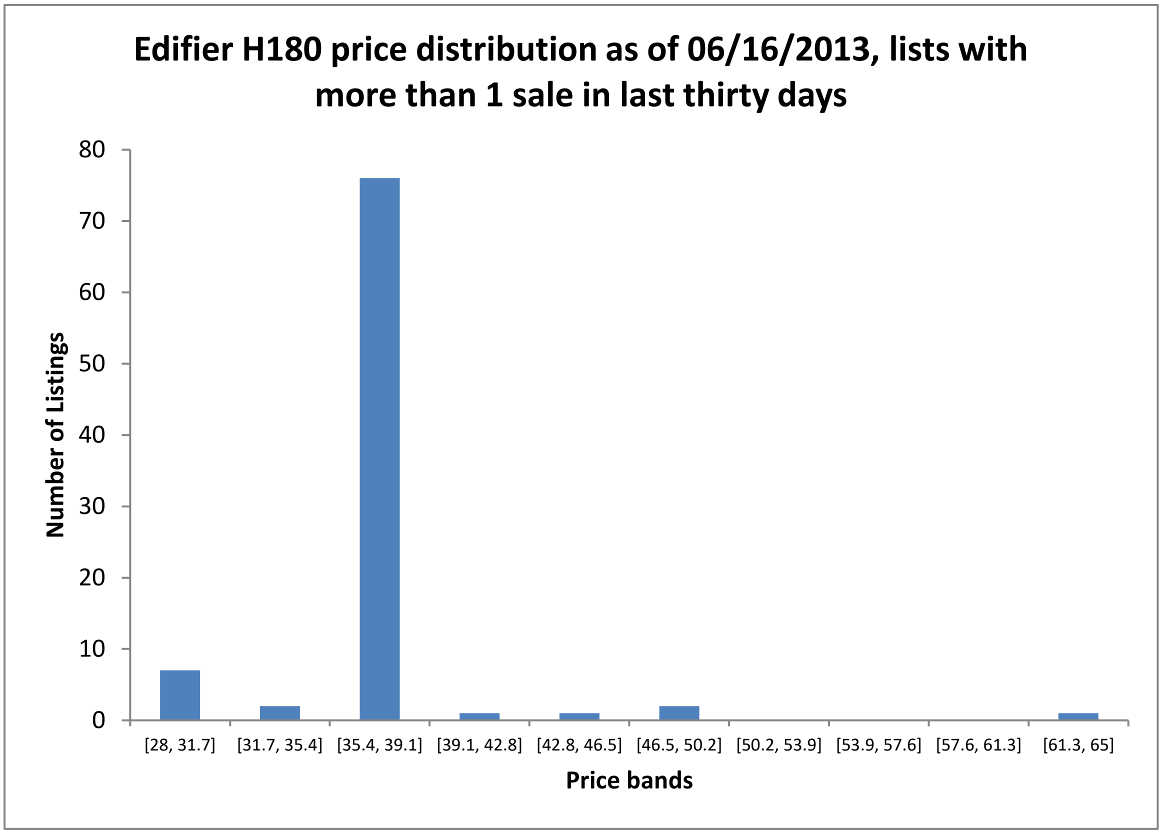


Figure 4.2 Edifier H180 price distribution as of 06/16/2013, taking out sellers with zero or one sales in last thirty days

The competition among the sellers with positive sales is stronger than the competition in the entire market. The minimum price charged by a seller having at least 2 sales in the last thirty days is ￥28 instead of ￥10. This indicates immediately the lists with the lowest prices do not have more than one sale in the last month. After taking out the sellers with zero or one sales in the last thirty days, the listing prices become more concentrated. 84.4% of all the lists have price range from

￥35.4 to ￥39.1, compared with 88.6% of the listing price range from ￥32.4 to ￥54.8 if sales information is not considered. The average price in Figure 4.2 is ￥38.7, with standard deviation of 4.33. The more concentrated price distribution and lower average price indicates the market is closer to perfect competition.

4.1.2 Case study on widely spread distribution

The other group includes top items from cameras, flash drives, laptops, cell phones and tablets. The market listing prices are widely spread. And, more importantly, the sellers distribute normally across price range as well.

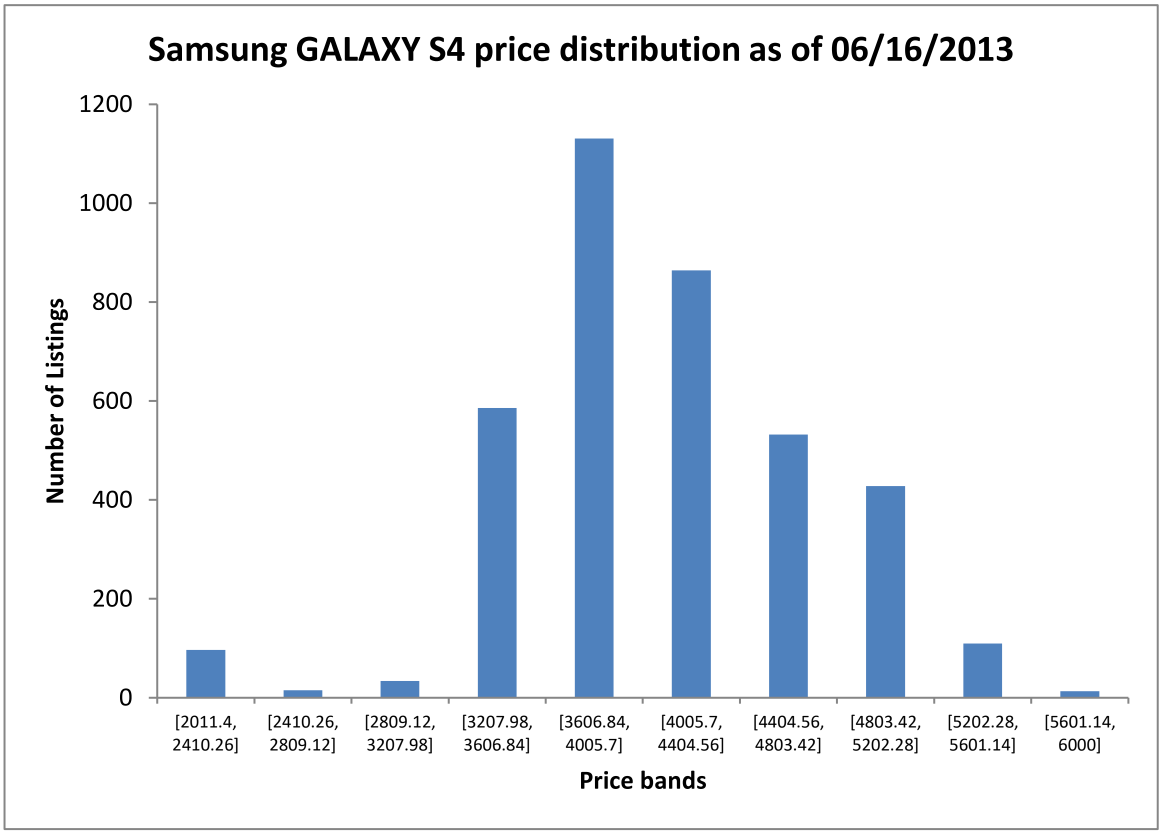


Figure 4.3 Samsung GALAXY S4 price distribution as of 06/16/2013

From the charts above we can see for the top ranked cell phone, sellers disagree on prices. The lowest listing price can be as low as ￥2011, while another seller could list it as high as ￥5601. The average price is ￥4077.5 with standard deviation of 717.6. Note there are not many sellers list below ￥3208. This indicates marginal cost for one cell phone may be around ￥3208. However, the cost information is not available in my research. Unless I have information about the marginal costs, I cannot conclude marginal revenue for sellers. The rational explanation for lists on the low-value end is hard.

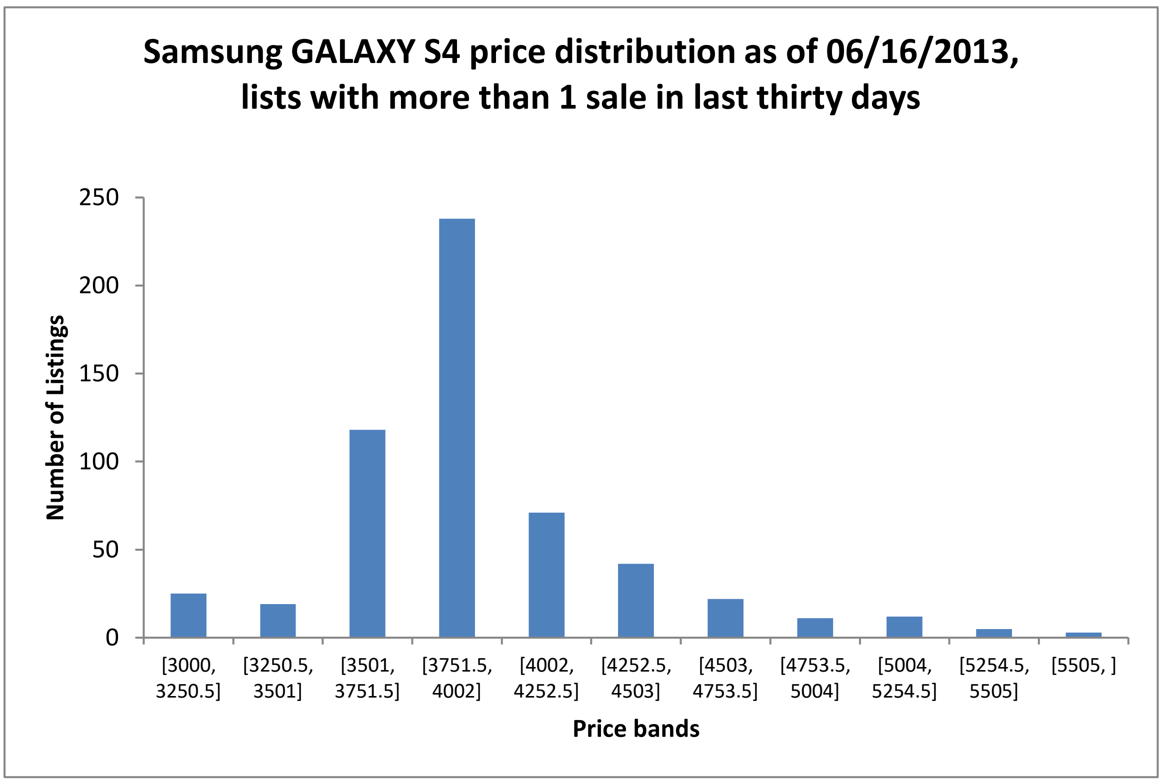


Figure 4.4 Samsung GALAXY S4 price distribution as of 06/16/2013, taking out sellers with zero or one sales in last thirty days

I examine the lists with two or more sales in last month again. This time, unlike the headphone example, the listing prices tend to normally distribute. The average price falls to ￥3634.3 with standard deviation of 611.6. The smaller standard deviation of the prices and lower average price indicates the market is closer to perfect competition. Although there is a large group of sellers, 41.9% of total, who have meaningful sales list price within [￥3752, ￥4002] group, there are noticeable amount of seller who get sales and list prices outside of that group. Especially, there are some sellers with sales even with prices above ￥4002, indicating sellers could survive with higher listing prices. Potentially this is led by the price differences in reputation and services. Also, the group of sellers who list price below ￥3000 nearly get no sales. This suggests that buyers are very cautious about the price. If the price is significantly lower than the median price, buyers may be suspicious of the listing.

4.1.3 The prices of the largest two sellers

I check the prices for the two sellers with the largest number of sales in the last thirty days. That sellers who do not have sales tend to price higher. But what if we compare the top sellers only with the sellers with sales? From the table below we can see among 93 items studied, for 49 the price listed by No. 1 seller is higher than the price listed by the No. 2 seller. And on average the prices listed by top two sellers are 95% and 92% of the median prices of all sellers with at least some sales in last 30 days.

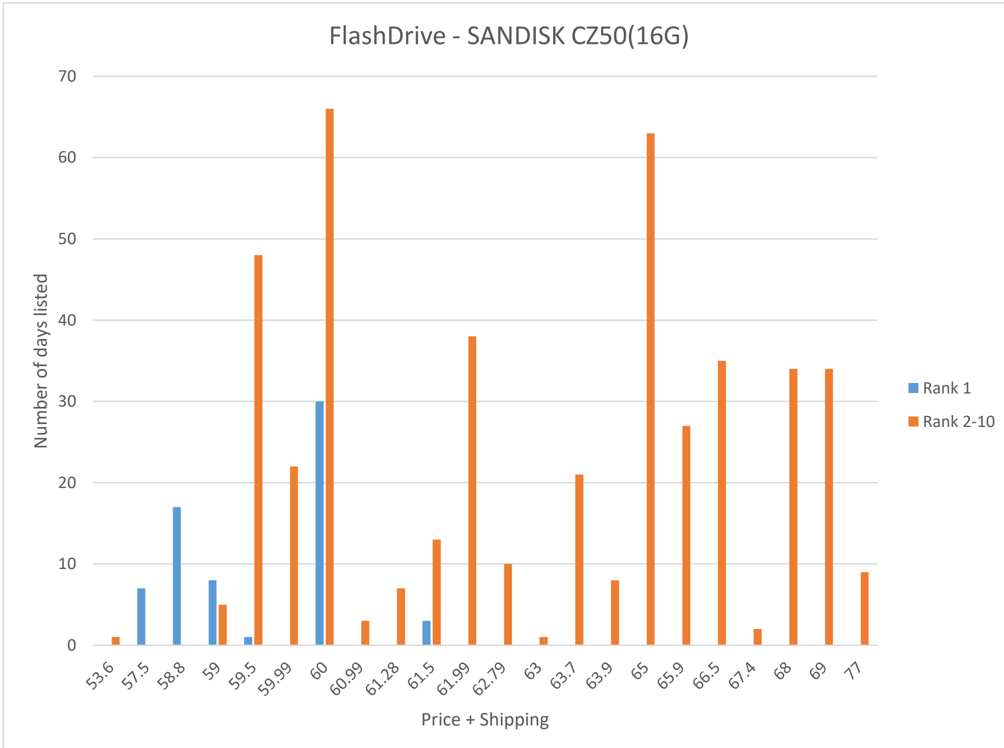
Table 4.1 Top two sellers' Prices on average by item

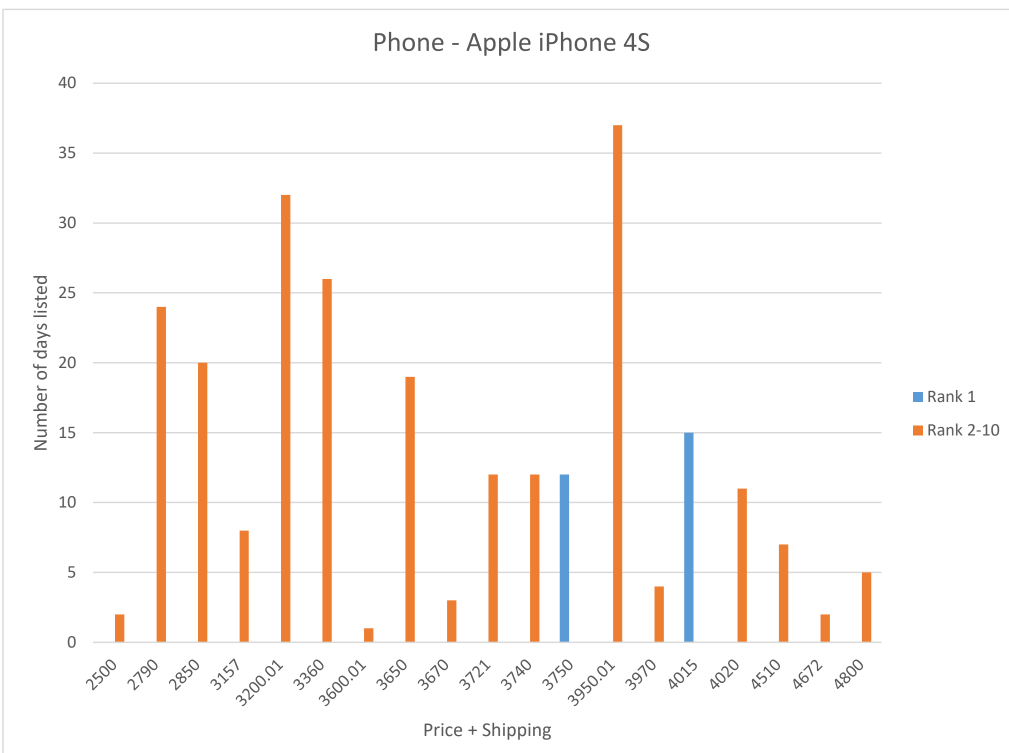
Category	Item Name	PriceShip1st	PriceShip2nd	PriceShip1st/ median	PriceShip2nd/ median
Camera	Canon IXUS 125 HS	905	886	91%	89%
	Canon IXUS 140	1,082	974	104%	93%
	Canon IXUS 240 HS	860	829	90%	86%
	Canon PowerShot A4000 IS	554	594	85%	92%
	Canon PowerShot D20	1,849	1,709	92%	85%
	Canon PowerShot SX500 IS	1,324	1,411	91%	97%
	Casio EX-N10	420	457	102%	111%
	Casio EX-TR150	5,119	5,248	96%	98%
	Casio TR200	5,417	4,864	104%	93%
	Nikon COOLPIX AW100s	1,375	1,322	92%	88%
	Samsung MV900F	866	876	80%	81%
	Sony DSC-W690	714	718	102%	103%
Flash Drive	AData S102(16G)	75	77	90%	91%
	Kingston DT101G2 (16GB)	60	58	90%	87%
	Kingston DT101G2 (32GB)	121	110	100%	91%
	Kingston DT101G2 (4G)	20	27	57%	77%
	Kingston DT101G2 (8G)	40	34	88%	75%
	Kingston DTI G3 (16G)	48	57	72%	86%
	Kingston DTI G3 (4G)	25	25	75%	75%
	Kingston DTI G3 (8G)	28	32	62%	70%
	Lenovo T180 (8G)	38	42	76%	86%
	PNY twin disk (8G)	43	46	95%	102%
	SANDISK CZ50(16G)	54	55	85%	86%
Headphone	AKG K420	277	286	100%	103%
	COGOO T02	36	36	100%	100%
	Edifier H180	39	39	100%	100%
	Edifier K550	43	52	100%	121%
	Electronic music DT-326	31	34	81%	88%
	Magic sound recording engineer Studio	1,076	1,041	67%	65%
	Meizu EP10	14	15	79%	85%
	MixStyle	21	25	99%	118%
	Philips SHP8000	276	306	92%	101%
	Sennheiser MX80	54	66	79%	97%
	Somic ST-2688	29	29	100%	100%
	Sony MDR-EX10A	48	55	83%	96%
Laptop	Acer E1-471G-53212G50Mn	2,565	2,394	84%	79%
	Asus A45EI323VD-SL	3,109	3,381	85%	92%
	Asus X401EI235A	2,292	1,880	82%	68%
	Asus X45EI237VD-SL	2,298	2,506	84%	91%
	Dell Ins15r-978	2,054	1,990	101%	98%
	Hasee K580S-I7 D0	3,933	3,888	94%	93%
	HP dv6-6029TX	3,473	3,096	106%	94%
	Lenovo G480-IFI	2,882	2,992	85%	88%
	Lenovo Y400N-IFI	5,036	4,711	102%	95%
	ThinkPad E430c(33651B8)	3,922	3,564	109%	99%

Mouse	ACER DS-1005	51	34	180%	121%
	Cherry JM-0300	100	95	101%	96%
	Dell MS111	28	21	98%	73%
	Lenovo M20	12	10	66%	56%
	Lenovo ThinkPad 57Y4635 black mouse	40	41	91%	92%
	Logitech G1	39	63	102%	165%
	Logitech G500	329	319	102%	99%
	Logitech M100	62	45	123%	90%
	Pennefather N6000	32	29	105%	96%
	Razer / Razer DeathAdder upgraded version	248	213	100%	86%
	Razer / Razer hell mad snake mirror version	149	110	122%	91%
Phone	Apple iPhone 4S	3,199	3,013	99%	93%
	Apple iPhone 5	4,565	4,365	106%	101%
	Daxian GS5000	82	79	95%	91%
	Daxian W111	67	77	86%	99%
	Huawei G520	704	694	121%	119%
	Lenovo A820T	731	787	121%	130%
	MIUI 2A(MI2A)	1,483	1,437	97%	94%
	MIUI 2S(MI2S)	1,906	1,857	98%	96%
	Nokia 1050	162	134	106%	88%
	Nokia 1120	37	32	40%	35%
	Nokia 2030	44	28	68%	43%
	Samsung GALAXY S4 I9500	3,075	3,112	85%	86%
	Samsung I9300 GALAXY SIII	2,105	2,112	87%	87%
Speaker	Cruiser R101T06	139	139	100%	100%
	Cruiser R10U	65	65	100%	100%
	Cruiser R201T08	189	189	100%	100%
	Dell AX210 USB	63	54	117%	100%
	Dell AX510	94	98	98%	102%
	Edifier R101V	119	119	100%	100%
	Edifier R151T	300	299	100%	100%
	JBL Duet	89	94	88%	93%
	Lenovo Lenovo L1520	37	34	81%	75%
	Microlab M-200 tenth anniversary edition	268	268	101%	101%
	Philips SPA1312	126	118	98%	92%
	Sound Pai think S020	31	33	74%	77%
Tablet	Apple iPad mini(16G)WIFI	2,320	2,074	105%	94%
	Apple iPad mini(32G)WIFI	2,689	2,658	96%	95%
	Apple iPad4(16G)WIFI	3,332	3,046	104%	95%
	Apple iPad4(32G)WIFI	3,628	3,586	98%	97%
	CUBE U25GT (8G) WIFI version	300	298	100%	99%
	Lenovo Ideatab A1000(4G)	799	841	100%	105%
	Lenovo Pad A1(16G)	297	504	31%	53%
	Samsung galaxy note 10.1N8000(16G)	3,245	1,722	120%	64%
	Samsung Galaxy Tab P3100(8G)3G	1,732	1,047	110%	66%
	Taipower P85 dual-core(16G)	498	518	99%	103%
	Taipower P88 quad-core (16G)	781	796	99%	101%
	window N70S (8G)WIFI	347	349	98%	98%
Average				**95%**	**92%**

4.1.4 Top seller's pricing strategy

There is a weak evidence of the firms adopting mixed pricing strategy described in Varian (1980), while some of the top sellers do not change their prices at all. To take a closer look at the top seller's pricing strategy, I plot the price distribution of the largest seller over the entire time window together with rank 2 to 10 sellers' prices. The charts below show the top sellers' price distributions for three categories. If the firms adopt the mixed pricing strategy, one should expect price distribution like in the first chart, where the top seller adjusts its price occasionally and the competitors do the same. However, there is no direct evidence showing the firms were randomly changing their prices. Moreover, example from the laptop market shows the case of the top seller do not change its price. This could be due to limited observing time, where the probability of a top seller changing its price multiple times is low in 77 days.


FlashDrive - SANDISK CZ50(16G)
Number of days listed
Price + Shipping
Rank 1
Rank 2-10
0
10
20
30
40
50
60
70
53.6
57.5
58.8
59
59.5
59.99
60
60.99
61.28
61.5
61.99
62.79
63
63.7
63.9
65
65.9
66.5
67.4
68
69
77


Phone - Apple iPhone 4S
Number of days listed
Price + Shipping
Rank 1
Rank 2-10
0
5
10
15
20
25
30
35
40
2500
2790
2850
3157
3200.01
3360
3600.01
3650
3670
3721
3740
3750
3950.01
3970
4015
4020
4510
4672
4800

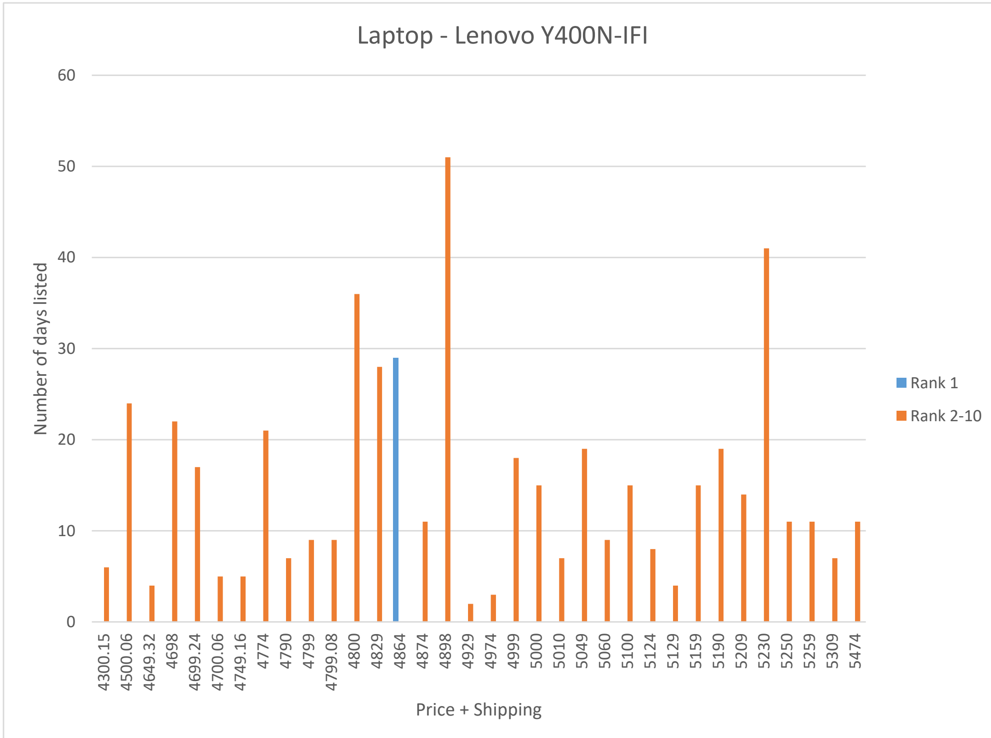


Figure 4.5 Examples of Pricing strategy of the top sellers

4.1.5 Price distribution of new lists and incumbents

The difference between the distribution of all listed prices and distribution of prices for only those listings by seller with sales in the last thirty days indicates that it is very misleading to draw conclusions based on the distribution of all listed prices. In fact, a large portion of these lists do not actually generate sales. Then a natural question to ask would be: are the new entrants pricing the same or at a lower level than incumbents? Intuitively, we might expect new entrants to list lower prices than incumbents because they have no sales histories and, very likely, thinner reputation histories than incumbents. However, it is also possible that lists with zero sales have higher prices than who have sales. If this is the case, the explanation would be that sellers who list higher prices do not get any customers. Since there are no charges for them to list, there is no reason for them to remove the list or lower the price if they currently do not want to engage. Sellers could also raise their price to avoid overselling. Think of the case when one seller runs out of its inventory. Any sellers can arbitrage this situation by raising the price to the lowest price available plus two shipping costs. One shipping cost covers the cost from buying from the lowest list. Another shipping cost covers the cost to ship to the buyer. There will be no additional loss other than shipping time. So, theoretically, sellers never have incentive to remove the lists. To evaluate this, I conduct t test for prices listed by sellers having at least one sale in last 30 days versus those who have zero sales.

Table 4.2 t test for prices of lists by seller with sales versus those without

ItemName	Category	Mean Price of lists with sales	Mean Price of lists without sales	t statistics	p value
Canon IXUS 125 HS	Camara	1,004.8	1,143.3	-37.7	0.00
Canon IXUS 140	Camara	1,001.1	1,115.9	-33.3	0.00
Canon IXUS 240 HS	Camara	1,010.1	1,251.6	-51.9	0.00
Canon PowerShot A4000 IS	Camara	664.5	813.5	-57.5	0.00
Canon PowerShot D20	Camara	1,933.8	1,912.7	1.7	0.08
Canon PowerShot SX50 HS	Camara	2,331.6	2,593.1	-50.8	0.00
Canon PowerShot SX500 IS	Camara	1,438.4	1,722.9	-66.0	0.00
Casio EX-N10	Camara	407.4	501.4	-27.4	0.00
Casio EX-TR150	Camara	5,473.3	5,594.2	-4.7	0.00
Casio TR200	Camara	5,248.3	5,439.8	-20.8	0.00
Nikon COOLPIX AW100s	Camara	1,513.0	1,908.1	-39.9	0.00
Samsung MV900F	Camara	1,077.5	1,358.7	-59.8	0.00
Sony DSC-W690	Camara	708.8	832.6	-45.2	0.00
AData S102(16G)	FlashDrive	87.1	92.4	-21.9	0.00
Kingston DT101G2 (16GB)	FlashDrive	69.0	80.7	-92.7	0.00
Kingston DT101G2 (32GB)	FlashDrive	121.9	149.0	-127.5	0.00
Kingston DT101G2 (4G)	FlashDrive	36.4	44.5	-55.0	0.00
Kingston DT101G2 (8G)	FlashDrive	46.5	54.6	-92.4	0.00
Kingston DTI G3 (16G)	FlashDrive	67.8	79.6	-109.2	0.00
Kingston DTI G3 (4G)	FlashDrive	34.0	45.3	-100.7	0.00
Kingston DTI G3 (8G)	FlashDrive	47.0	54.8	-51.9	0.00
Lenovo T180 (8G)	FlashDrive	52.1	72.5	-57.2	0.00
PNY twin disk (8G)	FlashDrive	48.8	63.8	-45.4	0.00
SANDISK CZ50(16G)	FlashDrive	66.6	78.8	-86.7	0.00
AKG K420	Headphone	263.9	275.9	-25.9	0.00
COGOO T02	Headphone	36.4	47.9	-40.6	0.00
Edifier H180	Headphone	39.2	45.2	-61.6	0.00
Edifier K550	Headphone	45.8	50.7	-58.5	0.00
Electronic music DT-326	Headphone	36.2	42.6	-25.4	0.00
Magic sound recording engineer Studio	Headphone	1,576.3	2,036.7	-65.6	0.00
Meizu EP10	Headphone	19.9	24.9	-42.4	0.00
MixStyle	Headphone	24.3	32.0	-35.9	0.00
Philips SHP8000	Headphone	308.3	431.9	-35.9	0.00

From the table above, we can see for all 94 items (see the rest in appendix), only 2 exhibit the phenomenon where prices listed by sellers who have zero sales in last 30 days are lower than the listing prices of sellers who had sales using a 5% level of significance. And the t test does not reject the null hypothesis that the two average prices are significantly different from zero. This indicates that sellers with positive sales generally list lower prices. There is no incentive for sellers with no sales to remove their lists because there is no cost to maintaining the list.

4.2 Price dispersion

4.2.1 Price dispersion measurement

Like the other price comparison websites, Taobao.com provides a possible way for consumers to obtain a list of price quotes for a given product at a nearly zero search cost. The clearing house model used in Baye, Morgan and Scholten (2004) seems to more closely match the environment that consumers encounter at price comparison sites. The distinguishing feature of a clearing house models is that identical firms sell to two types of consumers: Those who buy at the lowest price listed at the clearing house, and those who do not. Consumers who do not buy at the lowest listed price may be loyal to a particular firm or may be unwilling or unable to access the site. The model predicts that the level of price dispersion depends on the number of firms. The expected difference between the lowest two prices is greater in markets with a small number of firms than in the markets with a large number of firms. When the law of one price holds, all firms in the market charge the same price and these measures of price dispersion are all zero.

To measure price dispersion, there are multiple options. 1) the gap measurement used in Baye, Morgan and Scholten (2004), which is the difference between the lowest two prices 2) sales-weighted coefficient of variation 3) the interquartile range, which is the percentage difference between the 3rd quartile and 1st quartile of listing prices. All these measurements are used to examine whether price dispersion is a disequilibrium phenomenon that is being corrected over time. And they are all normalized to the same scale across different items.

I make two assumptions for simplicity. First, based on previous studies, I assume all buyers incorporate shipping cost into the final price, so all price dispersions measures are based on the final price: tag price plus shipping cost. Second, one may argue that the listing price may not represent the final price for each transaction, in some cases given there are different color and bundle options. This seems particularly true to the products which bundle options are important, such as cell phones. However, the listing prices I collect are the lowest prices that seller accepts for a basic package of the product. Although the add-on values of packages offered by different sellers can be different, the basic package, usually without any add-ons or accessories, should not differ significantly after controlling for the seller's reputation.

I first examine whether the gap measure is an effective measure of price dispersion. Aggregated level data are used to compare China's online market with results from the US online market studied by Baye, Morgan and Scholten (2004). That is, for each item, I will have only one observation, the difference between the two lowest prices listed by firms with positive sales for each day. Baye Morgan and Scholten (2004) find that no measure of dispersion is meaningful other than this gap measure. In their study, the price gap is defined as the difference between the lowest two prices. The classical Bertrand model implies that the gap between the two lowest prices should be zero in any equilibrium. However, the clearing house model predicts differently. Suppose the prices charged by $n \geq 2$ firms for a given product are ordered from lowest to highest, so that $p_1 \leq p_2 \leq \cdots \leq p_n$. Define "the gap", $G = p_2 - p_1$, to be the difference between the two lowest prices. The underlying assumption is that the two lowest price lists capture all shoppers. All buyers will only choose between the lowest two prices. Thus, G is a proxy for a quantity-weighted measure of price dispersion if, as seems likely, consumers visiting the price comparison site tend to be price-sensitive. We can also define the normalized price gap as

$$Price\ \ Gap\ \ between\ \ the\ \ two\ \ lowest\ \ prices\ \ (normalized) = \frac{The\ \ second\ \ lowest\ \ price - The\ \ lowest\ \ price}{The\ \ lowest\ \ price}$$

Since I have the sales data, I also calculate a sales weighted average price which is the sum of each list price multiplied by sales over the last thirty days divided by total sales by all firms over the last thirty days. The formula is

$$Sales\ \ weighted\ \ average\ \ price = \frac{\sum_{i=1}^{N}(Price + Shipping)_i \times Last\ \ 30\ \ day\ \ sales_i}{\sum_{i=1}^{N} Last\ \ 30\ \ day\ \ sales_i}$$

Where $N =$

$total\ \ number\ \ of\ \ firms\ \ listing\ \ a\ \ price\ \ for\ \ the\ \ item\ \ in\ \ that\ \ day$

The sales weighted average price can be used to calculate a sales weighted standard deviation of price uses similar method, which is calculated using the following formula. Non-zero count is defined as the total number of lists with at least 1 sale in last 30 days for a given day.

$$Sales\ \ weighted\ \ std.price = \sqrt{\frac{\sum_{i=1}^{N} w_i (Price_i - \overline{Price_w})^2}{(N' - 1) \sum_{i=1}^{N} w_i \big/ N'}}$$

Where $N =$

$total\ \ number\ \ of\ \ firms\ \ with\ \ postitive\ \ sales\ \ listing\ \ a\ \ price\ \ for\ \ the\ \ item\ \ in\ \ that\ \ day$

$$w_i = i^{th}\ \ list'sweight = \frac{Last\ \ 30\ \ day\ \ sales_i}{\sum_{i=1}^{N} Last\ \ 30\ \ day\ \ sales_i}$$

$N' =$

$total\ \ numberof\ \ lists\ \ with\ \ non\ \ zero\ \ sales\ \ in\ \ last\ \ 30\ \ days$

$$Price_i = i^{th}\ \ list's\ \ price + shipping$$

$$\overline{Price_w} = weighted\ \ average\ \ of\ \ (price + shipping)$$

To capture price dispersion across different items, I use the coefficient of variation, which is the sales weighted standard deviation divided by weighted average. The formula is listed below.

$$Sales\ \ weighted\ \ Coefficient\ \ of\ \ variation\ \ of\ \ price = \frac{Sales\ \ weighted\ \ standard\ \ deviation\ \ of\ \ price}{Sales\ \ weighted\ \ average\ \ price}$$

Last but not least, I adopt the interquartile range measurement and normalize it by dividing by the 1st quantile of price. This can be viewed as normalized range measurement. The formula is listed below.

$$Normalized\ \ quantile\ \ gap\ \ of\ \ price/shipping = \frac{3rd\ \ quantile\ \ of\ \ price/shipping - 1st\ \ quantile\ \ of\ \ price/shipping}{Average\ \ of\ \ price/shipping}$$

4.2.2 Dispersion over time

Price dispersion persists over time, regardless of the measurement method being used. The table below shows the average price dispersion measures described in the previous section by product. The average normalized price gap between the two lowest prices is 17.1%, which means the second lowest price is 17.1% higher than the lowest price on average. This measure can be as high as 102.4%, where the gap between the lowest prices is 102.4% of the lowest price. In market for Sennheiser MX80, the lowest price is ￥23.0 on average, while the second lowest price is￥47.3 on average. The average interquartile range (%) for all items is 27.2%, which means that on average the 75th percentile price is 27.2%, of the average price, higher than the 25th percentile. Most markets have reasonable price dispersion within 0% – 30% of the market average prices. Dispersion is smaller with the quantity of sales considered as demonstrated in section 4.1.2.

Table 4.3 Price dispersion measures by product

Category	Item Name	Average gap between the two lowest prices	Average nomalized interquatile	Average sales weighted coefficient of variation
Camera	Canon IXUS 125 HS	11.5%	23.3%	17.9%
	Canon IXUS 140	13.5%	17.6%	16.8%
	Canon IXUS 240 HS	7.0%	35.5%	26.8%
	Canon PowerShot A4000 IS	6.0%	35.2%	22.0%
	Canon PowerShot D20	25.5%	20.4%	36.5%
	Canon PowerShot SX500 IS	18.5%	23.6%	17.4%
	Casio EX-N10	22.7%	10.6%	15.3%
	Casio EX-TR150	38.4%	21.4%	18.7%
	Casio TR200	49.2%	12.7%	14.0%
	Nikon COOLPIX AW100s	17.2%	30.2%	23.7%
	Samsung MV900F	3.6%	38.8%	22.2%
	Sony DSC-W690	4.8%	15.7%	13.0%
Flash Drive	AData S102(16G)	17.7%	11.4%	9.5%
	Kingston DT101G2 (16GB)	22.0%	16.8%	9.7%
	Kingston DT101G2 (32GB)	30.0%	15.4%	11.9%
	Kingston DT101G2 (4G)	2.4%	37.5%	38.6%
	Kingston DT101G2 (8G)	26.7%	22.7%	16.6%
	Kingston DTI G3 (16G)	5.4%	16.0%	20.5%
	Kingston DTI G3 (4G)	7.9%	46.2%	31.1%
	Kingston DTI G3 (8G)	27.5%	24.8%	33.2%
	Lenovo T180 (8G)	3.4%	53.4%	34.4%
	PNY twin disk (8G)	3.1%	16.2%	8.7%
	SANDISK CZ50(16G)	8.3%	17.8%	11.7%
Headphone	AKG K420	4.4%	14.6%	8.1%
	COGOO T02	12.6%	25.2%	4.2%
	Edifier H180	21.0%	0.0%	7.0%
	Edifier K550	5.8%	8.2%	9.9%
	Electronic music DT-326	7.7%	46.4%	11.3%
	Magic sound recording engineer Studio	17.3%	60.5%	31.0%
	Meizu EP10	12.9%	57.0%	20.0%
	MixStyle	4.1%	50.9%	21.4%
	Philips SHP8000	4.3%	12.9%	5.7%
	Sennheiser MX80	102.4%	27.3%	14.6%
	Somic ST-2688	18.1%	16.8%	20.4%
	Sony MDR-EX10A	11.8%	25.1%	12.1%
Laptop	Acer E1-471G-53212G50Mn	8.4%	18.0%	20.1%
	Asus A45EI323VD-SL	9.5%	10.1%	11.7%
	Asus X401EI235A	50.3%	21.0%	19.0%
	Asus X45EI237VD-SL	20.0%	12.5%	13.9%
	Dell Ins15r-978	5.1%	18.5%	8.9%
	Hasee K580S-I7 D0	45.9%	11.3%	5.3%
	HP dv6-6029TX	17.8%	17.7%	9.1%
	Lenovo G480-IFI	2.9%	23.8%	19.4%
	Lenovo Y400N-IFI	7.4%	6.2%	5.6%
	ThinkPad E430c(33651B8)	15.4%	11.6%	8.1%
Mouse	ACER DS-1005	22.9%	49.7%	32.5%
	Cherry JM-0300	9.6%	0.0%	35.8%
	Dell MS111	26.5%	42.1%	32.7%
	Lenovo M20	14.1%	83.0%	130.8%
	Lenovo ThinkPad 57Y4635 black mouse	63.0%	33.3%	24.0%
	Logitech G1	17.3%	112.0%	63.7%
	Logitech G500	10.6%	27.6%	17.2%
	Logitech M100	6.0%	18.3%	17.6%
	Pennefather N6000	12.1%	75.1%	28.1%
	Razer / Razer DeathAdder upgraded version	3.5%	20.8%	12.3%
	Razer / Razer hell mad snake mirror version	10.4%	64.7%	14.7%
Phone	Apple iPhone 4S	2.7%	40.1%	23.0%
	Apple iPhone 5	1.5%	15.3%	15.3%
	Daxian GS5000	5.4%	32.1%	6.0%
	Daxian W111	17.4%	36.9%	12.8%
	Huawei G520	16.8%	51.7%	27.1%
	Lenovo A820T	6.0%	43.0%	21.0%
	MIUI 2A(MI2A)	42.4%	6.8%	5.4%
	MIUI 2S(MI2S)	11.5%	14.9%	8.1%
	Nokia 1050	4.6%	37.1%	30.6%
	Nokia 1120	10.7%	96.4%	98.0%
	Nokia 2030	1.5%	79.4%	62.6%
	Samsung GALAXY S4 I9500	4.5%	13.9%	14.6%
	Samsung I9300 GALAXY SIII	8.2%	25.6%	15.7%
Speaker	Cruiser R101T06	23.4%	0.0%	7.1%
	Cruiser R10U	38.7%	0.0%	8.3%
	Cruiser R201T08	100.7%	0.5%	11.7%
	Dell AX210 USB	20.0%	66.3%	28.6%
	Dell AX510	15.6%	30.6%	19.1%
	Edifier R101V	16.3%	0.0%	15.9%
	Edifier R151T	9.0%	0.0%	2.9%
	JBL Duet	5.0%	15.0%	26.5%
	Lenovo Lenovo L1520	8.3%	42.6%	38.1%
	Microlab M-200 tenth anniversary edition	41.7%	22.5%	11.7%
	Philips SPA1312	9.6%	14.7%	7.3%
	Sound Pai think S020	6.0%	31.2%	14.1%
Tablet	Apple iPad mini(16G)WIFI	11.0%	8.4%	9.0%
	Apple iPad mini(32G)WIFI	12.8%	9.1%	9.1%
	Apple iPad4(16G)WIFI	21.2%	6.3%	6.7%
	Apple iPad4(32G)WIFI	21.6%	5.8%	5.5%
	CUBE U25GT (8G) WIFI version	0.9%	7.2%	1.8%
	Lenovo Ideatab A1000(4G)	28.4%	1.1%	3.7%
	Lenovo Pad A1(16G)	6.0%	68.4%	52.0%
	Samsung galaxy note 10.1N8000(16G)	49.1%	66.8%	34.1%
	Samsung Galaxy Tab P3100(8G)3G	5.2%	34.0%	28.8%
	Taipower P85 dual-core(16G)	15.6%	7.8%	8.5%
	Taipower P88 quad-core (16G)	15.0%	4.1%	3.1%
	window N70S (8G)WIFI	1.7%	8.8%	7.2%
Average		**17.1%**	**27.2%**	**20.1%**

The average coefficient of variation weighted by sales is 0.201, meaning on average the standard deviation is 0.201 of the average price. This means on average the Chinese consumer electronic markets show some degrees of price dispersion. We can see that although price dispersion exists, more items have relatively mild dispersion rather than extreme dispersion. The chart below shows the trends of the three measures over time. In addition, I plot the average normalized price gap between the two lowest prices with positive sales. It is similar to the average normalized price gap between the two lowest prices, which means the dispersion is similar with and without sales information using a gap measure between the two lowest prices. There is no clear downward trend over time based on these measures. And the gap measure between the two lowest prices is clearly less stable than the other two measures.

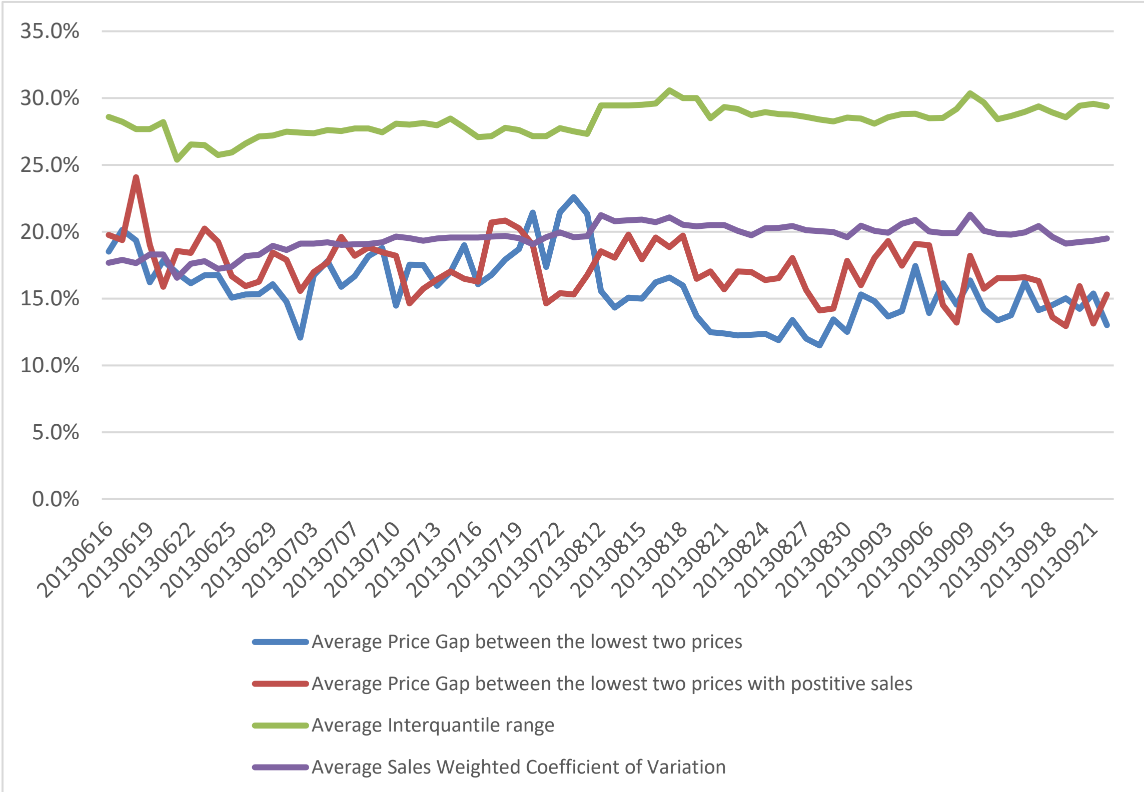


Figure 4.6 Average Price Gap between the largest two sellers, Average Price Gap between the largest two sellers with positive sales, Average Interquartile range and Average Coefficient of variation weighted by sales over time

4.2.3 Gap measurement between the two lowest prices

The gap measure between the lowest two listed prices does not mean much in the real world. Using the sales data, I calculate the market share of the two lists with the lowest prices for each item. The median of this market share is 0.03%. The highest value of this market share is 41.8% in one of the speaker markets. However, the third highest is 5.9% indicating that is a rare case. In nearly half of all the markets, the two lists with the lowest prices have zero sales.

The chance of the two lowest price lists are posted by the largest two sellers is low. Among the 93 items studied in the first day, for only 4 of them does one of the two lowest price lists rank in the top two in terms of quantity sold. They are Acer E1-471G-53212G50Mn (Laptop), CUBE U25GT (8G) WIFI version (Tablet), Lenovo T180 (8G) (Flash Drive) and Sound Pai think S020 (Speaker). The common attribute for these items is that the popularity of these items is not very high. They are cheap substitutes for similar products in some sense. Buyers who seek and potentially buy these items very likely to have sufficient knowledge of the quality of the product. When they try to purchase these items, they have a very high opportunity cost thus a significant low price would compensate the potential risk of the product.

4.3 Market competition and concentration

4.3.1 Number of competing firms

The data I collected from China' online electronic markets give an opportunity to study market concentration than existing studies on US online markets. In Baye, Morgan and Scholten (2004), they observe the consumer electronic market with several sub-markets but only limited number of firms in each sub-market. Among the 1,000 products they observed over eight months, around 6% of total the number of observations are single listings. Over 80% of the observations has listings from 30 firms or less. Observations with more than 55 listings represents less than 0.5% of the total. Simply speaking, the US consumer electronics markets they studied include many sub-markets with a few competitors in each. However, based on the data I collected from Taobao.com, the total number of lists, including listings both with and without sales, ranges from 55 to 17,675. The reason for number of lists vary dramatically from item to item could be supply restraint. Many of items listed on Taobao.com are imported. The sellers are more frequently located near the coast. Another important information is that number of lists with non-zero sale is highly correlated, with the total number of lists per day. The correlation coefficient between the total number of listings and the number of listings with no sales is 0.974.

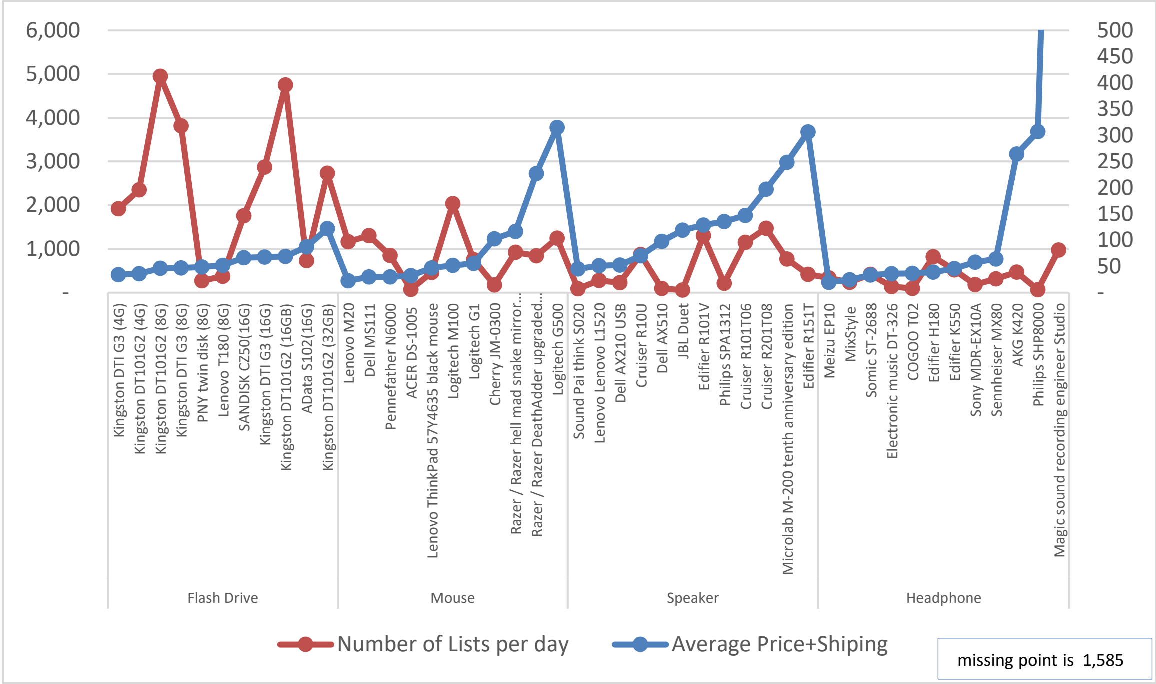
Table 4.4 Number of sellers with sales and without

						Correlation	0.974
Category	**Item Name**	**Number of Lists per day**	**Number of lists with non-zero sale per day**	**Category**	**Item Name**	**Number of Lists per day**	**Number of lists with non-zero sale per day**
Camera	Casio TR200	798	109	**Mouse**	Logitech M100	2,038	205
	Canon PowerShot SX500 IS	781	114		Dell MS111	1,303	336
	Canon IXUS 125 HS	717	59		Logitech G500	1,249	88
	Canon IXUS 240 HS	677	85		Lenovo M20	1,166	339
	Canon PowerShot A4000 IS	645	89		Razer / Razer hell mad snake mirror version	921	81
	Sony DSC-W690	549	68		Pennefather N6000	847	126
	Samsung MV900F	510	91		Razer / Razer DeathAdder upgraded version	846	79
	Canon IXUS 140	488	73		Logitech G1	755	110
	Casio EX-TR150	264	43		Lenovo ThinkPad 57Y4635 black mouse	465	129
	Nikon COOLPIX AW100s	259	28		Cherry JM-0300	179	55
	Canon PowerShot D20	212	44		ACER DS-1005	71	12
	Casio EX-N10	93	21	**Phone**	Apple iPhone 4S	17,675	1,881
Flash Drive	Kingston DT101G2 (8G)	4,950	627		Apple iPhone 5	17,197	1,975
	Kingston DT101G2 (16GB)	4,753	539		Samsung I9300 GALAXY SIII	10,664	1,229
	Kingston DTI G3 (8G)	3,813	438		Samsung GALAXY S4 I9500	6,397	1,112
	Kingston DTI G3 (16G)	2,871	299		Huawei G520	2,971	617
	Kingston DT101G2 (32GB)	2,731	233		MIUI 2S(MI2S)	2,605	417
	Kingston DT101G2 (4G)	2,348	288		Lenovo A820T	1,825	435
	Kingston DTI G3 (4G)	1,920	224		MIUI 2A(MI2A)	1,644	218
	SANDISK CZ50(16G)	1,758	182		Nokia 1050	1,108	371
	AData S102(16G)	735	81		Nokia 1120	964	356
	Lenovo T180 (8G)	374	66		Nokia 2030	805	296
	PNY twin disk (8G)	273	34		Daxian W111	297	111
Headphone	Magic sound recording engineer Studio	978	156		Daxian GS5000	149	45
	Edifier H180	819	151	**Speaker**	Cruiser R201T08	1,476	141
	Edifier K550	524	108		Edifier R101V	1,303	236
	AKG K420	473	99		Cruiser R101T06	1,151	107
	Somic ST-2688	417	86		Cruiser R10U	875	177
	Meizu EP10	336	104		Microlab M-200 tenth anniversary edition	768	60
	Sennheiser MX80	319	79		Edifier R151T	419	36
	MixStyle	233	44		Lenovo Lenovo L1520	283	36
	Sony MDR-EX10A	186	44		Dell AX210 USB	231	63
	Electronic music DT-326	139	27		Philips SPA1312	211	32
	COGOO T02	99	43		Dell AX510	100	31
	Philips SHP8000	67	8		Sound Pai think S020	89	17
Tablet	Apple iPad mini(16G)WIFI	6,706	844		JBL Duet	59	17
	Apple iPad4(16G)WIFI	5,644	695	**Laptop**	Lenovo Y400N-IFI	1,020	141
	Samsung galaxy note 10.1N8000(16G)	1,920	263		Acer E1-471G-53212G50Mn	889	61
	Apple iPad mini(32G)WIFI	1,822	166		Asus X45EI237VD-SL	732	103
	Apple iPad4(32G)WIFI	1,811	233		Asus A45EI323VD-SL	694	85
	Samsung Galaxy Tab P3100(8G)3G	1,657	206		Lenovo G480-IFI	575	60
	Taipower P85 dual-core(16G)	661	61		Asus X401EI235A	420	55
	Taipower P88 quad-core (16G)	454	55		Hasee K580S-I7 D0	134	34
	Lenovo Ideatab A1000(4G)	353	87		HP dv6-6029TX	66	24
	Lenovo Pad A1(16G)	320	23		ThinkPad E430c(33651B8)	66	14
	CUBE U25GT (8G) WIFI version	215	54		Dell Ins15r-978	55	15
	window N70S (8G)WIFI	201	58				

The dramatic difference of number of lists between Taobao.com and

Shopper.com could come from three reasons. First, the seller and buyer base is much larger because there is no charge for listing and transacting on Taobao.com. A simple fact is that Taobao's "Single day sale" volume is larger than the combined "Black Friday" sale amounts of eBay and Amazon. Shopper.com was not comparable with Taobao.com in terms of market size. Second, the internet is much more accessible and widely used for e-commerce today than in 2003. The latest numbers show in China the "Single day sale" volume has surpassed sales volume on the same day in physical stores. Last, Taobao.com is more likely to have a better system to classify product listings. As mentioned before, Taobao.com checks for incorrect listings, which ensures accuracy of the listing and prevents sellers from differentiating their products by naming slightly different.

While there is no strong linkage between item price and number of firms listing, it is worth noting that similar items have a similar number of lists. For example, in the flash drive category, Kingston DT101G2 (8G) and Kingston DT101G2 are the largest two sub-markets. The number of lists on these two markets per day are around 4,800 which is significantly higher than any other flash drives. Another example would be Apple iPhone 4S and Apple iPhone 5. They both have an average number of lists per day of 17,000, which is much higher than the third place Samsung I9300 GALAXY SIII. It is very likely that these lists have the similar sellers.



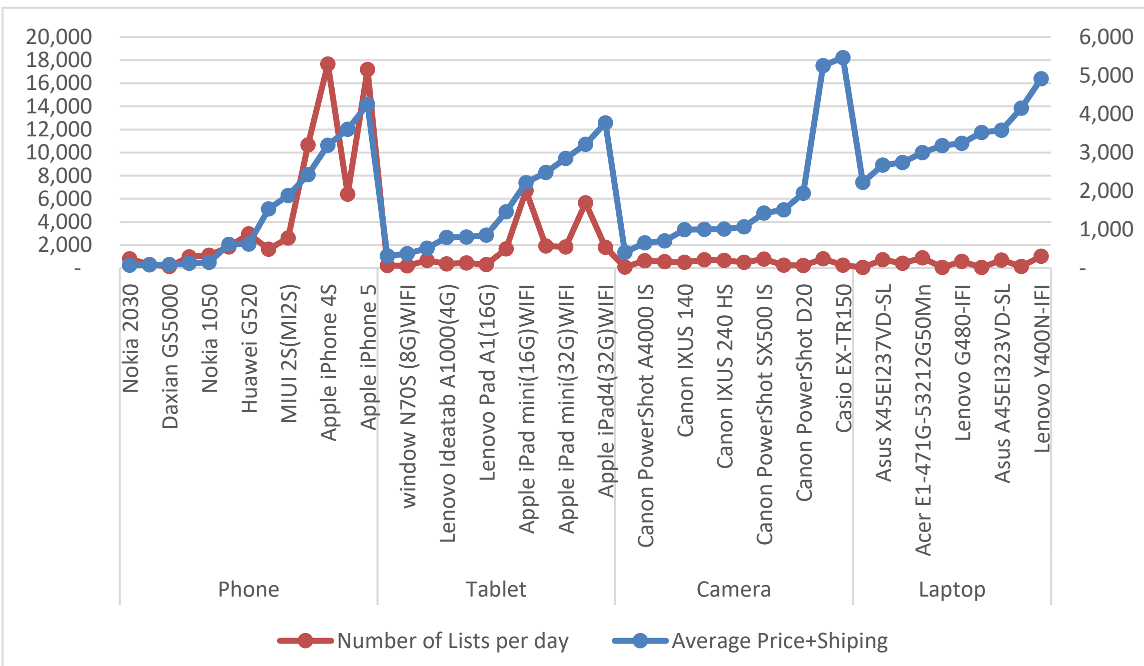


Figure 4.7 Number of lists per day of different categories, order by average price

Is there any linkage between product popularity and number of competitors? To see this, I look at the number of lists with different numbers of sales in last 30 days. From the table below we can clearly see that the number of sellers exponentially decreases as the number of sales go up. The number of sellers without any sales in flash drive market is 22,016 while the total number of lists is 25,246. The percentage of sellers without sales is 87.2%. Again, this proves that analysis on price dispersion without considering whether the list is actually selling or not will be misleading.

Table 4.5 Number of lists of different categories, order by last 30-day sale amount, as of June 16th, 2013

Last 30 day sales	Camera	Flash Drive	Headphone	Laptop	Mouse	Phone	Speaker	Tablet	Total
0	5530	22016	3417	4210	8635	15451	6122	14051	79432
1	406	1702	383	270	773	3861	406	1364	9165
2	171	472	128	115	271	1345	149	403	3054
3	72	242	78	48	113	645	71	201	1470
4	58	141	50	33	97	449	60	117	1005
5	27	90	37	27	57	291	31	71	631
6	32	65	20	11	43	206	29	62	468
7-9	53	129	48	40	83	441	45	130	969
10-14	51	89	46	30	81	359	32	120	808
15-19	46	64	33	11	35	215	24	53	481
20-29	37	68	29	26	40	278	23	83	584
30-49	40	61	25	18	42	227	23	73	509
50-99	48	38	39	16	32	250	18	69	510
100+	69	69	49	33	44	385	25	77	751

Total	**6640**	**25246**	**4382**	**4888**	**10346**	**24403**	**7058**	**16874**	**99837**

The last thing to examine is the relationship between average total number of sales in last 30 days and the average number of non-zero sale sellers. From the chart below, we can see the average total number of sales in last 30-day increases as the average number of non-zero sale sellers increases. Also, depending on the category, the relationships between the two are different. This indicates there are unobserved category characteristics.

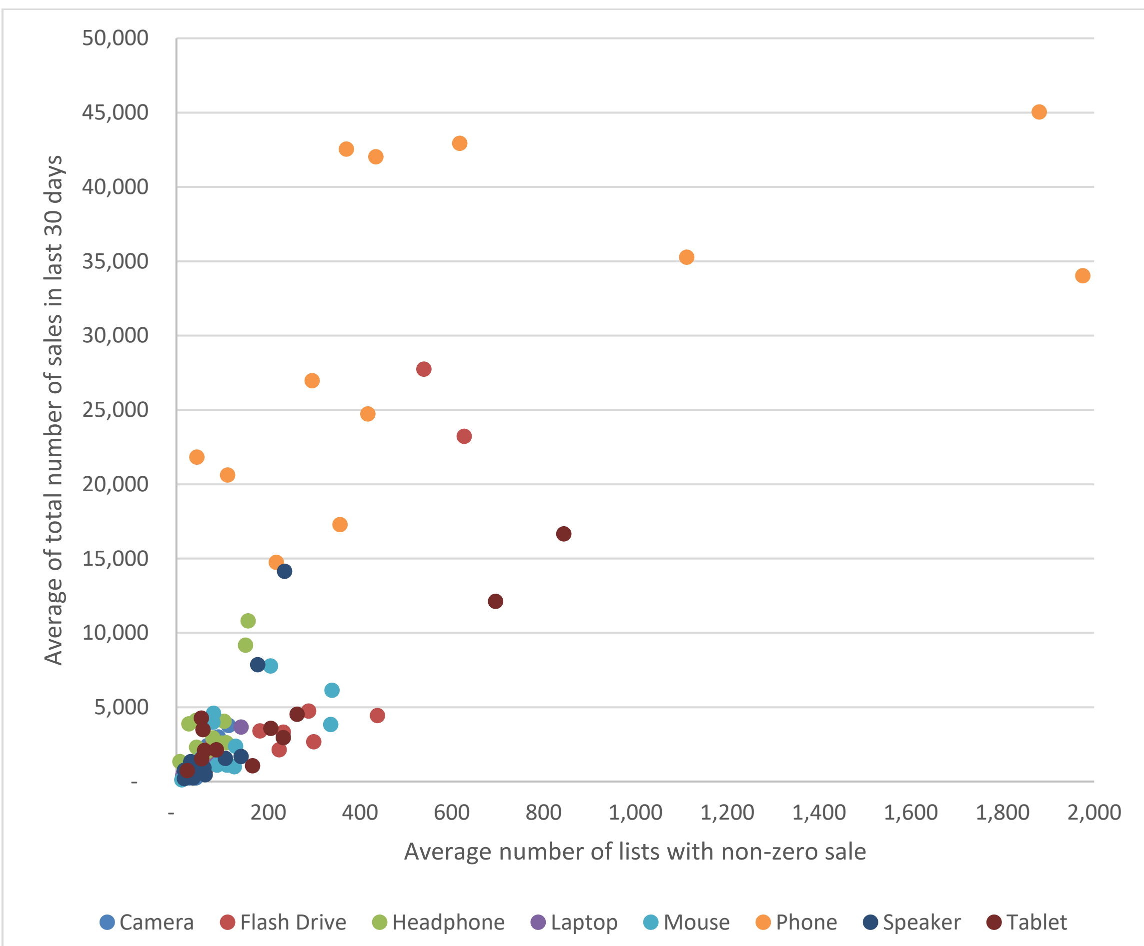


Figure 4.8 Scatter plot for average total number of sales in last 30 days and the average number of non-zero sale sellers by category

4.3.2 Market share of the largest two sellers

One of the key findings in Baye, Morgan and Scholten (2004) is that the price gap between the lowest two prices is highly correlated with the number of firms selling. It is interesting to see how dominating the top two sellers are and what prices that top two sellers list for sale. First, I look at market shares of the top two sellers for each item. In most cases, the largest two sellers are not dominating the market by themselves.

Table 4.6 Top two sellers' market shares on average by item

Category	Item Name	1st seller last 30-day sales	2nd seller last 30-day sales	Market share 1st seller	Market share 2nd seller	Market share difference	Market share top two sellers
Camera	Canon IXUS 125 HS	176	132	22%	16%	5%	38%
	Canon IXUS 140	854	225	35%	9%	26%	45%
	Canon IXUS 240 HS	376	250	23%	15%	8%	38%
	Canon PowerShot A4000 IS	349	282	22%	18%	4%	40%
	Canon PowerShot D20	219	152	23%	16%	7%	40%
	Canon PowerShot SX500 IS	832	539	22%	14%	8%	36%
	Casio EX-N10	100	39	35%	13%	21%	48%
	Casio EX-TR150	77	32	31%	13%	18%	44%
	Casio TR200	309	132	27%	11%	15%	38%
	Nikon COOLPIX AW100s	59	32	24%	13%	11%	37%
	Samsung MV900F	647	495	22%	17%	5%	38%
	Sony DSC-W690	786	350	33%	15%	18%	47%
Flash Drive	AData S102(16G)	1,590	345	60%	13%	47%	73%
	Kingston DT101G2 (16GB)	16,010	6,525	58%	24%	34%	81%
	Kingston DT101G2 (32GB)	1,670	705	50%	21%	29%	72%
	Kingston DT101G2 (4G)	1,456	790	31%	17%	14%	48%
	Kingston DT101G2 (8G)	12,573	2,452	54%	11%	44%	65%
	Kingston DTI G3 (16G)	449	248	17%	9%	8%	26%
	Kingston DTI G3 (4G)	481	340	23%	16%	7%	38%
	Kingston DTI G3 (8G)	960	505	22%	11%	10%	33%
	Lenovo T180 (8G)	263	136	30%	16%	14%	46%
	PNY twin disk (8G)	399	158	47%	18%	28%	65%
	SANDISK CZ50(16G)	2,323	162	68%	5%	64%	73%
Headphone	AKG K420	993	468	38%	18%	20%	57%
	COGOO T02	2,263	697	55%	17%	38%	72%
	Edifier H180	6,367	614	70%	7%	63%	76%
	Edifier K550	1,647	144	63%	6%	58%	69%
	Electronic music DT-326	3,616	158	93%	4%	89%	97%
	Magic sound recording engineer Studio	1,420	1,172	13%	11%	2%	24%
	Meizu EP10	1,766	632	44%	16%	28%	59%
	MixStyle	575	223	44%	17%	27%	62%
	Philips SHP8000	937	369	70%	27%	42%	97%
	Sennheiser MX80	1,133	612	39%	21%	18%	60%
	Somic ST-2688	587	298	35%	18%	17%	53%
	Sony MDR-EX10A	1,483	372	64%	16%	48%	80%

Laptop	Acer E1-471G-53212G50Mn	137	91	22%	15%	8%	37%
	Asus A45EI323VD-SL	314	217	27%	19%	8%	46%
	Asus X401EI235A	282	161	41%	23%	17%	64%
	Asus X45EI237VD-SL	289	200	22%	15%	7%	37%
	Dell Ins15r-978	346	75	69%	15%	54%	84%
	Hasee K580S-I7 D0	594	98	59%	10%	49%	68%
	HP dv6-6029TX	212	90	43%	18%	25%	61%
	Lenovo G480-IFI	234	137	27%	16%	11%	42%
	Lenovo Y400N-IFI	996	426	27%	12%	16%	39%
	ThinkPad E430c(33651B8)	316	23	81%	6%	75%	87%
Mouse	ACER DS-1005	38	22	38%	21%	16%	59%
	Cherry JM-0300	762	194	48%	12%	36%	60%
	Dell MS111	1,270	344	33%	9%	24%	42%
	Lenovo M20	1,192	754	19%	12%	7%	32%
	Lenovo ThinkPad 57Y4635 black mouse	789	416	33%	18%	16%	51%
	Logitech G1	228	158	21%	14%	6%	35%
	Logitech G500	402	125	36%	11%	25%	48%
	Logitech M100	5,211	726	67%	9%	58%	76%
	Pennefather N6000	270	132	27%	13%	14%	41%
	Razer / Razer DeathAdder upgraded version	2,686	306	68%	8%	60%	75%
	Razer / Razer hell mad snake mirror version	3,930	259	86%	6%	80%	92%
Phone	Apple iPhone 4S	2,402	1,814	5%	4%	1%	9%
	Apple iPhone 5	3,699	2,715	11%	8%	3%	19%
	Daxian GS5000	14,845	3,193	68%	15%	53%	83%
	Daxian W111	11,079	5,566	54%	27%	27%	81%
	Huawei G520	4,079	3,257	10%	8%	2%	17%
	Lenovo A820T	6,420	3,759	15%	9%	6%	24%
	MIUI 2A(MI2A)	3,974	2,396	27%	16%	11%	43%
	MIUI 2S(MI2S)	5,706	4,643	23%	19%	4%	42%
	Nokia 1050	9,245	4,198	22%	10%	12%	32%
	Nokia 1120	3,974	2,270	23%	13%	10%	36%
	Nokia 2030	6,007	2,957	22%	11%	11%	33%
	Samsung GALAXY S4 I9500	5,542	3,404	16%	10%	6%	25%
	Samsung I9300 GALAXY SIII	4,032	3,095	9%	7%	2%	15%
Speaker	Cruiser R101T06	617	396	40%	25%	14%	65%
	Cruiser R10U	5,179	617	66%	8%	58%	74%
	Cruiser R201T08	735	354	44%	21%	23%	65%
	Dell AX210 USB	76	64	17%	14%	3%	32%
	Dell AX510	115	88	27%	21%	6%	48%
	Edifier R101V	10,600	1,169	75%	8%	67%	83%
	Edifier R151T	340	102	62%	19%	43%	81%
	JBL Duet	90	31	45%	15%	29%	60%
	Lenovo Lenovo L1520	57	32	25%	14%	11%	39%
	Microlab M-200 tenth anniversary edition	474	153	52%	17%	35%	69%
	Philips SPA1312	1,115	66	84%	5%	79%	89%
	Sound Pai think S020	486	170	65%	23%	42%	87%
Tablet	Apple iPad mini(16G)WIFI	4,263	2,140	26%	13%	13%	38%
	Apple iPad mini(32G)WIFI	256	126	24%	12%	12%	36%
	Apple iPad4(16G)WIFI	2,798	1,181	23%	10%	13%	33%
	Apple iPad4(32G)WIFI	531	390	18%	13%	5%	31%
	CUBE U25GT (8G) WIFI version	2,883	457	68%	11%	57%	78%
	Lenovo Ideatab A1000(4G)	665	447	31%	21%	10%	52%
	Lenovo Pad A1(16G)	408	230	55%	31%	24%	86%
	Samsung galaxy note 10.1N8000(16G)	726	357	16%	8%	8%	24%
	Samsung Galaxy Tab P3100(8G)3G	969	291	27%	8%	19%	35%
	Taipower P85 dual-core(16G)	920	241	44%	12%	33%	56%
	Taipower P88 quad-core (16G)	758	189	49%	12%	37%	62%
	window N70S (8G)WIFI	1,800	700	51%	20%	31%	72%
Average				**39%**	**14%**	**25%**	**53%**

The table above shows quantity sold and the market of the top two seller of each product. The sellers are ranked by the total quantity sold through the three-month period. The largest market share for a top list is 93% in the Electronic music DT-326 (headphone) market, where the market average daily last-30-day sale is 3,873 and the top seller has an average daily last-30-day sale of 3,616. It is also the market where the top two sellers take the largest share. The lowest market share for a top seller is only 5% in the iPhone 4s market. Also, it is also the market with the smallest share by the top two sellers. The largest market share for second largest seller is 31% in the Lenovo Pad A1 4G (tablet) market. Overall, the average market share for the top seller is 39% and for the second largest seller is 14%. The average market share difference between the top two sellers is 25%. And the average combined market share of top two sellers is 53%.

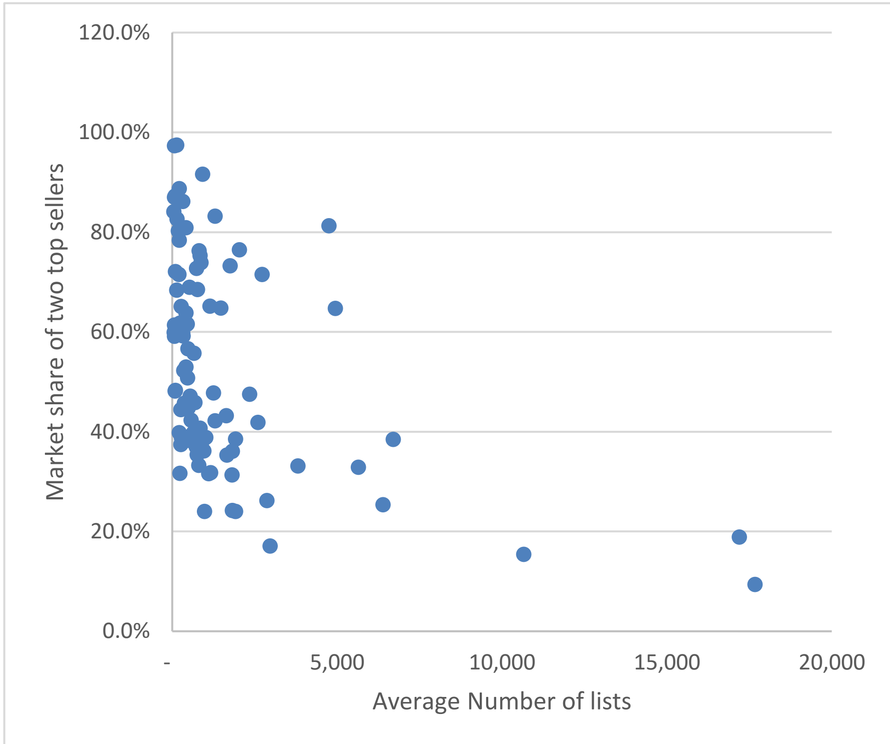


Figure 4.9 Top two sellers' market shares against average number of lists

There is a weak linkage between the number of sellers and the market share of the largest two sellers for cameras. The chart above shows the average market share of the largest two sellers and average number of sellers per day. There is a moderate negative relationship between the number of lists and market share of the top two sellers. This negative relationship is even stronger in the cell phone and tablet markets than for headphones and laptops. As the average number of lists increases, the market share of the top two sellers decreases quicker. The econometric estimation of the relationship is presented in the next section 4.3.3.

4.3.3 Market share of the top ten sellers

The market share of the top 10 sellers is high, and it also shows a very strong negative relationship against the number of listings. The market share of the top 10 sellers for each item. The market share for the top 10 sellers, except in three of the markets, are all above 50% with an average of 80.6%. Considering the number of sellers is typically in the hundreds, this strongly suggest these markets are dominated by a few large sellers. The figure below shows the negative relationship between market size and market share of the top sellers. The correlation is high as -0.67. This means as the market size increases, the market share of the large sellers decreases correspondingly.

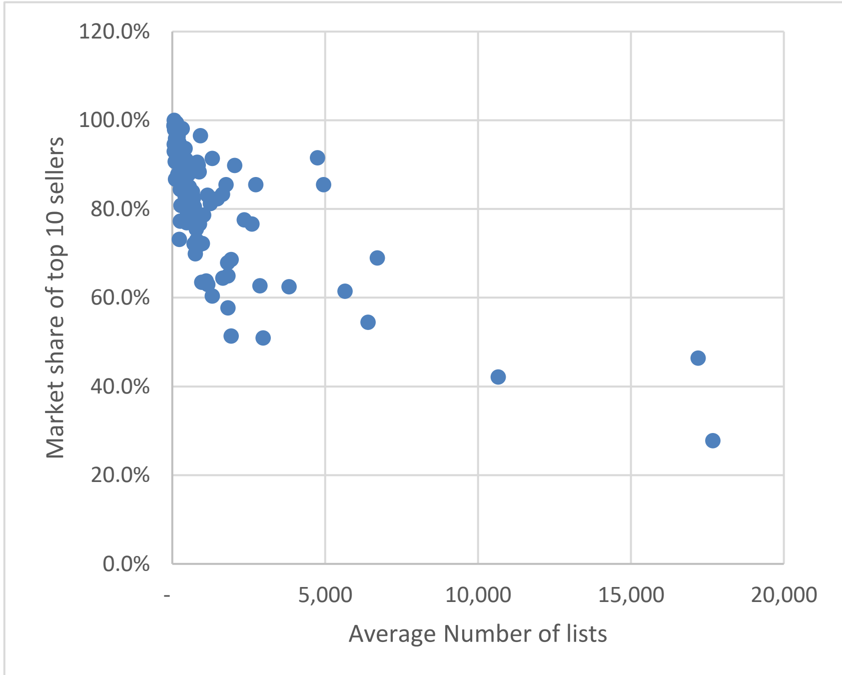


Figure 4.10 Market share of top 10 sellers against the average number of lists

The comparison of the characteristics of market share of top 2 sellers and of top 10 sellers show that the markets studied were captured by the top 10 sellers in each market instead of only the top 2 sellers. The average of the market share of the top 2 sellers is 0.532, while the average of the market share of the top 10 sellers is 0.806. The standard deviation for the market share of the top 10 sellers is 0.143, which is also smaller than the standard deviation for the market share of the top 2 sellers, 0.210. The maximum of the market share of the top 2 sellers is similar to the maximum of the market share of the top 10 sellers. They are both close to 1.

However, the minimum of the market share of the top 2 sellers is smaller the minimum of the market share of the top 10 sellers. In the 93 items studied, the smallest market share of the top 10 sellers is 0.278. This indicates that the markets studied were majorly captured by the top 10 sellers in each market.

Table 4.7 The characteristics for market share of top 2 sellers and of top 10 sellers

	Market share of top 2 sellers	Market share of top 10 sellers
Average	0.532	0.806
Standard Deviation	0.210	0.143
Minimum	0.094	0.278
Maximum	0.974	1.000

I ran OLS regression to evaluate the impact from the number of lists on the market share of top 2 sellers, as well as on the market share of top 10 sellers. The coefficients for number of listings are negative in both regressions. The market share for the top sellers fall as the number of competing firms increases. It follows the economic intuition. The impact of competition is more pronounced on the market share of the top two sellers than on the market share of the top ten sellers.

Table 4.8 Market share of top 2 sellers and of top 10 sellers against the average number of lists

Dependent variable	***Market share of top 2 sellers***	***Market share of top 10 sellers***
Intercept	0.580*	0.826*
Number of lists	-3.109E-05*	-1.311E-05*
R^2	0.1847	0.0704

*** significant under 5% level of significance**

Sellers also dynamically change their pricing strategy in reaction to change in item popularity. It is difficult to measure this factor directly, but the total number of sales should be a good proxy. This is particularly true for consumer electronic product markets with rapid new technology growth. I notice the prices for some items increased over the period examined and for other items prices declined. This might be caused by competition among items within the category i.e. substitution effects between similar products. When the item is popular, sellers may be able to strategically increase their lists to maximize profits. As an item becomes less popular, the price is expected to decrease in the long term. For example, in the data there are listings for both the Apple iPhone 4s and Apple iPhone 5. When Apple launched iPhone 5, iPhone 4s sales are expected to drop. On the other hand, as one item becomes more popular, new sellers could enter the market and strategically price lower to gain market share. In this case, more popular means more competitors and more possibility of price competition. More popularity can lead price to diverge more than usual. The table below shows the correlation between sales weighted coefficient of variation and total number sold in the last 30 days by category. Except for headphones and the computer mouse category, all other categories show negative correlation which means prices diverge more with larger number of total sales. The headphone market shows little correlation while the computer mouse market shows relative strong positive correlation. This indicates the computer mouse market is very different from the others. One or a few sellers dominate the market and there is less price dispersion with a larger number of total sales.

Table 4.9 Correlation between sales weighted coefficient of variation and total number sold in the last 30 days by category

Category	**Correlation**
Camera	-0.228
Flash Drive	-0.269
Headphone	0.035
Laptop	-0.364
Mouse	0.242
Phone	-0.076
Speaker	-0.151
Tablet	-0.172

4.4 Dispersion-competition model

4.4.1 Gap measurement between the lowest two prices

In this section I adopt the model of Baye, Morgan and Scholten (2004) in order to test whether relationship between the gap measure and the number of firms is similar in Chinese and U.S. online markets. Although the price gap between the lowest two lists shows weak relationship with market structure, I want to see if the China online market data show the same result as the US online market. Baye, Morgan and Scholten found that in price dispersion is greater in the market with small number of firms. They concluded that price competition increases as the number of firms in the market increases using the difference between the lowest two listing prices as the measure of price dispersion. They argued that the gap measure is only meaningful measure for price dispersion. In their study, the sales information was not available. Follow their assumption, I use the price gap between the lowest two lists and the number of sellers on the market to run regression. To count for the potential fixed effects, I also run the regressions with controls for category difference and item difference. Model 1 is the baseline model without control for any fixed effects. The only regressor is the number of lists. Model 2 adds the total quantity sold in the last thirty days to test the impact from the market size. Model 3 controls for the fixed effects by category. Model 4 controls for the fixed effects by individual product. The formula for the models are

Model 1:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices_{kt} = \alpha + N_{kt} + \varepsilon$$

Model 2:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices_{kt}$$
$$= \alpha + N_{kt} + Q_{kt} + \varepsilon$$

Model 3:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices_{kt}$$
$$= \alpha_c + N_{kt} + Q_{kt} + \varepsilon$$

Model 4:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices_{kt}$$
$$= \alpha_k + N_{kt} + Q_{kt} + \varepsilon$$

Where α is the common intercept, α_c are the fixed effects for category, α_k are the fixed effects for individual product, N_{kt} is the number of firms in market k at time t, Q_{kt} is the total quantity sold in last 30 days for market k at time t,ε is the common error term.

Visually, the gap measurement against the market size shows a weak negative relationship. In Figure 4.11, the horizontal axis is the number of lists in each market.

The vertical axis represents the normalized price gap between the lowest two listings. Most standardized price gap are close to zero.

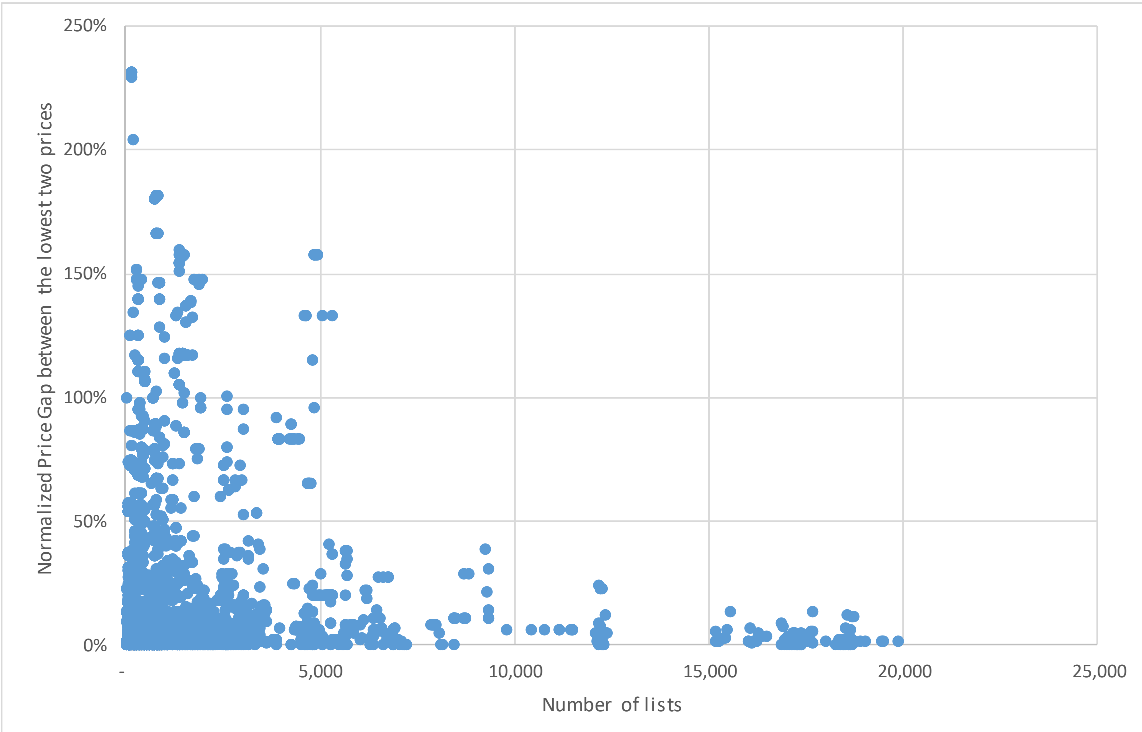


Figure 4.11 Normalized Price gap against Number of lists

The finding from Baye, Morgan and Scholten (2004) can be reproduced using China's online data if fixed effects are not controlled. Table 4.10 summarizes the OLS regression results for the four models. Model 1 is the regression without fixed effects, which is used to compared with the results in Baye, Morgan and Scholten (2004). The coefficient for the number of firms is negative and statistical significant.

But the regression only has an R^2 equals to 0.010, which means the variance explained by the number of firms is very small. Although the direction of the relationship is consistent with the findings from Baye, Morgan and Scholten (2004), the reason is very likely to be that two random small sellers charging the lowest prices are more likely to appear in a large market.

Table 4.10 Price-competition models with price gap between the lowest two prices

Dependent variable	***Normalized Gap between the lowest two prices***			
	Model 1	*Model 2*	*Model 3*	*Model 4*
Intercept	1.695e-01*	1.732e-01*	1.728e-01*	1.670e-01*
Number of lists	-7.21E-05*	-5.13E-05*	-3.14E-06	-6.66E-06
Total sales of the market in last 30 days		-1.08E-06*	5.15E-07	-1.65E-07
fixed effects				
Category	N	N	Y	N
Individual product	N	N	N	Y
R^2	0.010	0.013	0.052	0.254

*** significant under 5% level of significance**

Model 2 shows that the total sales of the market in last 30 days shows negative impact on price gap, when he fixed effects are not controlled. Given the same number of firms in the market, a larger market size, as indicated by the quantity sold, drives smaller price gap between the lowest two prices. This means in a more popular market the dispersion is smaller, when the fix effects are not controlled.

Model 3 and model 4 show that, controlled for category or item fix effects, the impact from the number of lists become insignificant. This means the price difference between the lowest two prices cannot be explained by the change of number of sellers within each market. The number of lists can only explain the price gap variance between markets. More firms are associated with small price gap between the lowest two prices. But controlled for fixed effects, the impact from the number of firms goes away, meaning the more firms does not drive smaller gaps.

In Baye, Morgan and Scholten (2004), the reason they use the lowest two sellers' prices as price dispersion measurement is that sales data is not available. They simply assumed that the market is captured by only these two sellers. My results show there is aggregation effects between the markets for the gap measure. And the impact disappears at disaggregated level. There are systemic differences among the categories or products on average, instead of within each category or product. This means that the gap measure is not an effective measure for price dispersion.

There may be some new sellers coming into the market and try to list unreasonable low price to attract buyers. So, I eliminate the sellers without any sales and redo the regressions above. The formula for the models are

Model 1:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices\ with\ positive\ sales_{kt} = \alpha + N_{kt} + \varepsilon$$

Model 2:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices\ with\ positive\ sales_{kt} = \alpha + N_{kt} + Q_{kt} + \varepsilon$$

Model 3:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices\ with\ positive\ sales_{kt} = \alpha_c + N_{kt} + Q_{kt} + \varepsilon$$

Model 4:

$$Normalized\ price\ gap\ between\ the\ two\ lowest\ prices\ with\ positive\ sales_{kt} = \alpha_k + N_{kt} + Q_{kt} + \varepsilon$$

Where α is the common intercept, α_c are the fixed effects for category, α_k are the fixed effects for individual product, N_{kt} is the number of firms with positive sales in market k at time t, Q_{kt} is the total quantity sold in last 30 days for market k at time t,ε is the common error term.

The results are shown in the table below. The parameter estimates change slightly but the direction and significance level do not change much. The conclusion is similar. The consistent estimate of the market structure makes the interpretation of the relationship economically less meaningful. In previous section, I already showed the market share of the lists with the lowest two prices. They typically capture a large portion of the market but price items near the market median price. The measurement between the lowest pricing sellers has little value in explaining the price distribution of the market.

Table 4.11 Price-competition models with price gap between the lowest two prices with positive sales

Dependent variable	***Normalized Gap between the lowest two prices with positive sales***			
	Model 1	*Model 2*	*Model 3*	*Model 4*
Intercept	1.869e-01*	1.896e-01*	1.773e-01*	0.1670*
Number of lists with positive sales	-6.988e-05*	-1.804e-05*	4.08E-06	-4.03E-06
Total sales of the market in last 30 days		-1.737e-06*	-6.54E-07	-9.83E-08
fixed effects				
Category	N	N	Y	N
Individual product	N	N	N	Y
R^2	0.005	0.008	0.018	0.371

*** significant under 5% level of significance**

4.4.2 Normalized interquartile measurement

In this section I test alternative price dispersion measure, normalized interquartile. In Baye, Morgan and Scholten (2004), the price difference between the highest and the lowest is also considered as a measure for price dispersion. But like coefficient of variation, it was not found to correlated negatively with the number of firms. By observing the actual data, I find the price range, the price difference between the highest and the lowest, may not be stable due to the large number of sellers. Even Taobao.com regularly checks for listings with extreme prices, some sellers still post lists with unreasonable prices. This could cause noise for my analysis. Therefore, I use interquartile to measure price dispersion and normalize it by the average price of the product. I run similar regressions as in precious section but replacing the measure for price dispersion by normalized interquartile. To count for the potential fixed effects, I also run the regressions with controls for category difference and item difference. Model 1 is the baseline model without control for any fixed effects. The only regressor is the number of lists. Model 2 adds the total quantity sold in the last thirty days to test the impact from the market size. Model 3 controls for the fixed effects by category. Model 4 controls for the fixed effects by individual product. The formula for the models are

Model 1:

$$Normalized\ \ interquartile_{kt} = \alpha + N_{kt} + \varepsilon$$

Model 2:

$$Normalized\ \ interquartile_{kt} = \alpha + N_{kt} + Q_{kt} + \varepsilon$$

Model 3:

$$Normalized\ \ interquartile_{kt} = \alpha_c + N_{kt} + Q_{kt} + \varepsilon$$

Model 4:

$$Normalized\ \ interquartile_{kt} = \alpha_k + N_{kt} + Q_{kt} + \varepsilon$$

Where α is the common intercept, α_c are the fixed effects for category, α_k are the fixed effects for individual product, N_{kt} is the number of firms in market k at time t, Q_{kt} is the total quantity sold in last 30 days for market k at time t,ε is the common error term.

Table 4.12 Price-competition models with normalized interquartile

Dependent variable				***Normalized price interquartile***
	<u>*Model 1*</u>	<u>*Model 2*</u>	<u>*Model 3*</u>	<u>*Model 4*</u>
Intercept	2.673e-01*	2.632e-01*	2.507e-01*	7.181e-02*
Number of lists with positive sales	3.514e-05*	-4.246e-05*	-6.933e-05*	3.005e-04*
Total sales of the market in last 30 days		2.601e-06*	-1.331e-06*	5.867e-07*
fixed effects				
Category	N	N	Y	N
Individual product	N	N	N	Y
R^2	0.003	0.015	0.136	0.964

*** significant under 5% level of significance**

The impact of the number of lists on price inter quartile is different at overall level and at disaggregated level. Model 1 is the regression without fixed effects. The coefficient for the number of firms is positive and statistical significant. But the regression only has an R^2 equals to 0.003, which means the variance explained by the number of firms is very small. The coefficient for the number of firms becomes negative and statistical significant in model 2. This shows the impact of the number of firms is different whether the market size is controlled or not. Model 2 shows that the total sales of the market in last 30 days shows positive impact on price gap, when the fixed effects are not controlled. Given the same number of firms in the market, a larger market size, as indicated by the quantity sold, drives larger price interquartile.

This means in a more popular market the dispersion is larger, when the fix effects are not controlled.

Model 3 shows that, controlled for category fix effects, both the impact from the number of lists and the total sales of the market in last 30 days are negative. This means the price interquartile decreases as the number of sellers increases within the product category. The result is different from using price gap as the price dispersion measure. However, it is also inconsistent with the result without control for category fixed effects. This mean the relationship between the number of lists and the price interquartile is different at aggregated level and the category level.

Model 4 shows that, controlled for individual product fix effects, both the impact from the number of lists and the total sales of the market in last 30 days are positive. This means the price interquartile increase as the number of sellers increases within each product. The result is different from the result without control for fixed effects, showed in model 2. This mean the relationship between the number of lists and the price interquartile is different at aggregated level and the individual product level.

4.4.3 Weighted coefficient of variation measurement

Using sales weighted coefficient of variation as the price dispersion measurement, I run four OLS regressions. Model 1 does not control for any fixed effects. The only regressor is the number of lists. Model 2 adds the total quantity sold in the last thirty days to test the impact from the market size. Model 3 controls for the fixed effects by category. Model 4 controls for the fixed effects by individual product. The formula for the models are

Model 1:

$$\ln(Sales\ \ weighted\ \ coefficient\ \ of\ \ variation)_{kt} = \alpha + N_{kt} + \varepsilon$$

Model 2:

$$\ln(Sales\ \ weighted\ \ coefficient\ \ of\ \ variation)_{kt} = \alpha + N_{kt} + Q_{kt} + \varepsilon$$

Model 3:

$$\ln(Sales\ \ weighted\ \ coefficient\ \ of\ \ variation)_{kt} = \alpha_c + N_{kt} + Q_{kt} + \varepsilon$$

Model 4:

$$\ln(Sales\ \ weighted\ \ coefficient\ \ of\ \ variation)_{kt} = \alpha_k + N_{kt} + Q_{kt} + \varepsilon$$

Where α is the common intercept, α_c are the fixed effects for category, α_k are the fixed effects for individual product, N_{kt} is the number of firms in market k at time t, Q_{kt} is the total quantity sold in last 30 days for market k at time t,ε is the common error term.

Table 4.13 Price-competition models with sales weighted coefficient of variation

Dependent variable	***ln(Sales weighted coefficient of variation)***			
	<u>Model 1</u>	*<u>Model 2</u>*	*<u>Model 3</u>*	*<u>Model 4</u>*
Intercept	-1.950e+00*	-1.946e+00*	-1.726e+00*	-3.630e+00*
Number of lists with positive sales	3.358e-04*	4.025e-04*	2.698e-04*	5.617e-04*
Total sales of the market in last 30 days		-2.234e-06*	-3.535e-06*	3.79E-07
fixed effects				
Category	N	N	Y	N
Individual product	N	N	N	Y
R^2	0.019	0.019	0.190	0.820

*** significant under 5% level of significance**

The market competition indicator, number of sellers with non-zero sales, is statistically significant in all four models. The positive coefficients indicate that competition effect does not reduce dispersion. This means the price dispersion measurement, considered the quantity of sales, is larger in markets in which there are more listings with positive sales. Recall that Baye, Morgan and Scholten (2004) fine that price dispersion is larger in markets with a small number of sellers than in those with a large number of sellers. The empirical result using China's online electronic data shows the opposite is true when we incorporate sales data. The price dispersion is smaller in the larger markets in terms of sales but not number of competitors. The sales weighted coefficient of variation measurement indicates that increasing the number of sellers in the market leads to larger dispersion.

There are multiple explanations for increasing price dispersion as the number of firms increases. In a more competitive market, sellers can differentiate from others by denoting more efforts on listing information and accessories. Ellison and Ellison (2009) find online sellers intend to frustrate consumer search by making the lists more complicated, causing consumers to be less price sensitivity. Remember all the prices studied are the lowest price for the item and it is assumed that the price only for base model option. However, buyers can browse and compare bundles since it barely has cost associated other than time. That is, the base model for a particular seller maybe higher than another. But a buyer may find it sells lower for a specific bundle that the buyer is looking for. Research on bundling strategy requests more detailed information of exact cost of the accessories. It is a good topic for future research.

Total sales of the market in last 30 days has a negative impact on price dispersion, as shown in model 2 and model 3. When the individual product fixed effects are controlled, the impact of the market size becomes insignificant. This means at overall level, given the same number of firms in the market, a larger market size, as indicated by the total quantity recently sold, drives larger price dispersion. This means in a more popular market the dispersion is larger, when the individual product fix effects are not controlled.

Are the results above robust? To see whether the relationship holds in all categories, I re-run Model 3 by category. In previous section, I discussed the different characteristics for different categories. The general relationship may not be enough to explain the true relationships within each category. The price-concentration model by category is specified as below.

$$ln(WCoV_{kt}) = \alpha_l + N_{kt} + Q_{kt} + \varepsilon$$

where $WCoV_{kt}$ is the sales weighted coefficient of variation for item *k* at time *t*; α_l is the constant intercept for category *l*; N_{kt} is total number of non-zero sales lists for item *k* at time *t*; Q_{kt} is the weekly total sales for item *k* at time *t*; ε is the common error term.

Table 4.14 Price-competition models with sales weighted coefficient of variation, by category

Dependent variable	***ln(Sales weighted Coefficient of variation)***					
Category	Intercept		Number of list with non-zero sales		Total sales of the market in last 30 days	
Camera	-1.49E+00	*	-2.69E-04		-5.55E-05	*
FlashDrive	-1.97E+00	*	2.18E-04	*	-4.51E-05	
Headphone	-2.48E+00	*	1.10E-03	*	-3.75E-05	
Laptop	-2.34E+00	*	1.27E-03	*	-4.84E-04	*
Mouse	-1.13E+00	*	-3.42E-04		4.63E-05	*
Phone	-1.93E+00	*	-7.40E-06	*	7.68E-06	*
Speaker	-1.85E+00	*	-5.16E-04		3.26E-06	*
Tablet	-2.51E+00	*	3.76E-04	*	-1.48E-04	*
R^2			0.2787			

*** significant under 5% level of significance**

The story from estimation by category is consistent with regard to the number of lists with non-zero sales. The coefficient on this parameter is positive or not significant away from zero in all categories, except phones. Again, this contradicts the results found by Baye, Morgan and Scholten. Within categories the sign of the coefficients on the total sales are not consistent with the overall market results. For computer mouse, cell phones and speakers, larger amount of sales is associated with larger price dispersion. For cameras, laptops and tablets, larger amount of sales is linked with smaller price dispersion.

4.5 Dispersion-concentration model

Somewhat counterintuitively, the results above indicate that price dipersion actuall increases with the number of firms. To further explore this relationship I now incorporate a measure of market concentration in the analysis. I use the Herfindahl index as the market concentration indicator. I run two additional OLS regressions on top of the dispersion-competition models. Model 1 adds the Herfindahl index, with number of competitors with positive sales as another regressor. Model 2 includes total quantity sold in the last thirty days to test the impact of total market size. The formula for the models are

Model 1:

$$\ln(Sales\ \ weighted\ \ coefficient\ \ of\ \ variation)_{kt} = \alpha + N_{kt} + H_{kt} + \varepsilon$$

Model 2:

$$\ln(Sales\ \ weighted\ \ coefficient\ \ of\ \ variation)_{kt} = \alpha + N_{kt} + Q_{kt} + H_{kt} + \varepsilon$$

Where α is the common intercept, α_c are the fixed effects for category, α_k are the fixed effects for individual product, N_{kt} is the number of firms in market k at time t, H_{kt} is the Herfindahl index in market k at time t, Q_{kt} is the total quantity sold in the last 30 days for market k at time t, and ε is the common error term.

Table 4.15 Price-concentration models

Dependent variable	***ln(Sales weighted coefficient of variation)***	
	Model 1	*Model 2*
Intercept	-1.406e+00*	-1.406e+00*
Number of lists with positive sales	-3.290e-05	-2.213e-05
Herfindahl index	-1.945e-04*	-1.944e-04*
Total sales of the market in last 30 days		-3.558e-07
fixed effects		
Category	N	N
Individual product	N	N
R^2	0.240	0.240

*** significant under 5% level of significance**

The Herfindahl index coefficient is negative and statistically significant in both Model 1 and Model 2. Remember the Herfindahl index equals to 10,000 when there is only one firm in the market and drops when there are more firms sharing the market. Both total sales and the number of sellers with non-zero sales are statistically insignificant. The results imply that as market concentration increases, as indicated by a higher value of the Herfindahl index, price dispersion actually decreases. This is consistent with the counter intuitive result from section 4.3.3. Factors typically associated with increased competition – either an increase in the number of firms listing a price or a reduction in the Herfindahl index – actually lead to greater price dispersion.

4.6 Price-reputation model

In order to further test how individual seller reputation impacts pricing strategy, a listing level model is developed. I test two versions of price dispersion for individual listings: the deviation from the mean price and the deviation from the top seller's price. Presumably, sellers do not collect whole market data like me. Instead, they use the weekly average price posted on the website as a benchmark. Recently there are services provided by Taobao.com to help 3rd party sellers grow their business. This is a sign of growing use of econometric methods in real world business. Back to 2013, Taobao.com did not have such a service. The formulas are

$$Price\ \ deviation\ \ from\ \ mean\ \ price = \frac{Price\ \ of\ \ seller\ \ i - weekly\ \ average\ \ price\ \ posted}{weekly\ \ average\ \ price\ \ posted}$$

$$Price\ \ deviation\ \ from\ \ top\ \ seller\ \ price = \frac{Price\ \ of\ \ seller\ \ i - price\ \ listed\ \ by\ \ top\ \ seller}{price\ \ listed\ \ by\ \ top\ \ seller}$$

Taobao.com provides a number of fields describing individual seller's reputation but not all of them have good data quality. Among the 1,198,356 records of individual lists, 1,062,577 do not have lifetime credit information. Lifetime credit information is the total number of feedback since first day the seller started listing prices on Taobao.com. The mechanism originates from eBay and has been gradually phased out from Taobao.com. Taobao.com then started to use a so-called “dynamic rating system”, where buyers can find the relative rating compared to “similar” sellers. Unlike the US market, Chinese online sellers usually get very high ratings. Thus taobao.com decided to present the buyers a comparable rating rather than absolute numbers. It could be biased since a similar seller may not sell the same item. However, given the data constraint, I will use the relative feedback rating system information to approximate the seller’s feedback rating. The total number of observations used is 253,359 after omitting missing records.

The additional services provided by sellers, described below, are controlled by dummy variables. All these dummies are expected to have a positive sign in the regression. Tmall is the trademark to identify a B2C seller. Warranty means the seller is willing to provide service after the purchase. Donation means the seller will donate a portion of its income from the item. Deposit means the seller has some deposit at Taobao.com to pay for potential disputes. Here are descriptive statistics of the individual reputation information and dummies.

Table 4.16 Variables for individual seller reputation and services

	Min	25% percentile	Median	Mean	75% percentile	Max
DescripRelativeRate	-100	0	14.69	27.29	46.26	100
AttitudeRelativeRate	-100	-0.51	8.9	24.8	40.91	100
SpeedRelativeRate	-100	0	14.2	26.85	44.34	100
Tmall	0	0	0	46.6%	1	1
Warranty	0	0	0	2.6%	0	1
Donation	0	0	0	0.1%	0	1
Pay as it arrives	0	0	0	0.6%	0	1
3 times refund if fake	0	0	0	3.5%	0	1
7 Day Return	0	0	1	71.8%	1	1
Imported	0	0	0	0.4%	0	1
Authorized dealer	0	0	0	0.2%	0	1
Deposit	0	0	0	42.8%	1	1
Credit Card Payable	0	0	0	8.9%	0	1

Three dynamic reputation ratings are the information Taobao.com displays to buyers. The ratings compare each current seller to its competitors. However, how Taobao.com determines competitors is unclear. It is very likely that Taobao.com compare the current seller with all sellers in the same category. The ratings have positive and negative numbers. It is more efficient than the absolute rating system since usually Chinese customers tend to rate the sellers very close to the best. And the shopping experience shows that dynamic ratings are the first piece of information displayed when buyers try to check seller reputation. For example, it takes several clicks to find how long the seller has been active on Taobao.com. However the information about this firm's dynamic rating is enabled by scripts hence it is able to be seen when point to the seller's name on the listing page. Notice the mean and median are both higher than zero. It is because the sellers selected are "better" sellers who have sales in the last month.

Service variables are used to control for additional cost and product differentiation. Among all services, three of them apply to most sellers. They are B2C seller, 7-day return service and deposit. Four services have very low usage among sellers. They are donation, pay as it arrives, certified imported product and authorized seller. All of them have less than 1% usage by all sellers.

To examine the relationship between price deviation and reputation, a fixed effects model can be applied, where a constant is estimated for each item. In this way, the unobserved market characteristics are first minimized. The key variable is the number of sellers with non-zero sales. Although it is proven to be significant in the aggregate level model, the effect may be muted once individual reputation information is added. The fixed effects price dispersion-reputation regression functional form is:

$$PD_{mt} = \alpha_k + MS_{mt} + Reputation_{mt} + Services_{mt} + \varepsilon$$

where PD_{kt} is the price deviation for individual seller *m* at time *t*; α_k is the k^{th} item specific intercept; MS_{mt} is the market share for m^{th} seller at time *t*; $Reputation_{mt}$ and $Services_{mt}$ are controlled for reputation and services provided by the seller; ε is the common error term.

Table 4.17 Price-reputation models with price deviation from average price

Dependent variable	***Percentage difference from average price***		
	Model 1	*Model 2*	*Model 3*
Intercept	-0.1307*	-0.1252*	-0.2579*
Market share	-0.2699*	-0.2476*	-0.2620*
Reputation			
Descriptive Relative Rate	0.0013*	0.0012*	0.0019*
Attitude Relative Rate	-0.0002*	-0.0002*	-0.0006*
Speed Relative Rate	0.0001*	0.0002*	0.0003*
Services			
Tmall	0.0217*	0.2648*	0.2896*
Warranty	0.1648*	0.1991*	0.2099*
Donation	-0.1409*	-0.1582*	-0.1394
Pay as it arrives	-0.0280*	-0.0484*	-0.0398
3 times refund if fake	0.1132*	0.0930*	0.1096*
7 Day Return	-0.0261*	-0.0234*	-0.0236*
Imported	0.1031*	0.0876*	0.1065*
Authorized dealer	0.1425*	0.1864*	0.1976*
Deposit	0.1049*	0.1247*	0.1159*
Credit Card Payable	-0.0173*	-0.0292*	-0.0175*
fixed effects			

Category	N	Y	N
Individual product	N	N	Y
R^2	0.038	0.061	0.112

*** significant under 5% level of significance**

Model 1 is the baseline model with no fixed effects. Model 2 and model 3 control fixed effects for category and individual product, respectively. While controlling for reputation, market share has negative impact on price. Reputation variables have little impact. All other things the same, a seller with a better reputation relative to its competitors is expected to take advantage and charge a higher price. However, the empirical results show that among the three types of relative reputation variables, attitude rating has a negative sign. This means the prices are loser for sellers with better attitude ratings. It is possible that buyers may not clearly understand the meaning of this rating. Compared to the other two, easy to understand ratings, attitude seems ambiguous. In most of the cases, buyers do not contact sellers before purchasing. So some buyers may interpret attitude as a rating for pricing. More importantly, the coefficients associated with reputation are very small. The largest increment from reputation is a 1% increase relative to competitors in description results leads to a 0.0019% higher price. If a seller is the top among competitors, with a rating 100% higher than others, it will price 0.19% higher than the average price. The payback from higher reputation is negligible. This is probably because the tendency of Chinese buyers to rate all sellers very highly, meaning reputation provides very little useful information.

Additional services do not always bring additional pricing advantage. Out of 10 services estimated, 4 have negative signs. A negative coefficient means if the seller does not have this service, it prices higher than those who do. The reasons vary. For two of them, donation and pay as it arrives, the percentage of seller having them are very small. So most buyers do not usually see these services and may not pay attention to it. Donation does not add value on price because buyers may think the seller put the additional cost into the price. Buyers tend to choose items without donation thus there is downward pressure on pricing. Pay as it arrives is not an attractive service because Taobao.com provides a third party secure transacting system like PayPal. Buyers can pay several days after receiving the product instead of paying right after receiving it. Similarly, the credit card payable service is not an important service because the 3rd party payment system is automatically registered when buyers register their account on Taobao.com. Interestingly, the other two services also having very little popularity, imported and authorized dealers, have a positive impact on pricing. This indicates buyers value these two services when shopping. Another service does not have positive influence on pricing is 7 days return service. A 7 day return service means buyers can return purchased items with any reasons. The explanation that a large portion of sellers have this service, 71.8%. So the sellers who do not have this service must have other services not controlled to maintain their position in the market. Typically, these sellers price higher and not having this service links with higher prices.

In order to better estimate the impact of reputation and services, I further investigate the price difference between the top sellers using difference in difference model. This method is useful when there are unobserved characteristics. Taking the difference against the top seller eliminates the unobserved difference between markets. This also eliminates the common market attributes such as number of competing firms. For the top 10 sellers with reputation information in each market, I take the percentage difference between the listing price and the listing price of the top seller. Similarly, for all the reputation and service dummies, I take the absolute difference. This will eliminate the unobserved information in the specific market and generate unbiased results. The difference in difference price dispersion-reputation regression functional form is:

$$\%D(P_{mt}) = D(MS_{mt}) + D(Reputation_{mt}) + D(Services_{mt}) + \varepsilon$$

where $\%D(P_{kt})$ is the percentage price deviation from the top seller for the top 10 individual seller *m* with reputation information at time *t*;$D(\ MS_{mt})$ is the market share difference from the top seller for *m*th seller at time *t*; $D(Reputation_{mt})$ and $D(Services_{mt})$ are difference from the top seller of reputation and services provided by the seller; ε is the common error term.

Table 4.18 Price-reputation difference in difference models

Dependent variable	***Percentage difference from top seller's price***	
	Model 1	*Model 2*
Intercept		
Difference in Market share	-0.1380*	-0.2204*
Reputation		
Difference in Descriptive Relative Rate	0.0060*	0.0054*
Difference in Attitude Relative Rate[1]	-0.0021*	-0.0006
Difference in Speed Relative Rate	-0.0021*	-0.0016*
Services		
Difference in Tmall	0.4381*	0.2630*
Difference in Warranty	0.1038*	0.3429*
Difference in Imported	0.0178	0.3038*
Difference in Authorized dealer	0.3360*	0.1921*
Number of sellers used		
	Top 2	Top 10
R^2	0.187	0.107

*** significant under 5% level of significance**

Model 1 includes only the second largest sellers, while model 2 includes the top 10 sellers in each market. The results from difference in difference models for top sellers are similar to the pooled fixed effects models. The table above shows the coefficient estimates. The difference in difference approach explains more variance in price dispersion, evidenced from the higher R-square. The key variable of interest, Market share, remains significantly negative. This means the top sellers with larger market share tend to list lower prices. After dropping the low frequency service dummies, almost all the remaining service variables turn significant. As all four difference in services dummies are positive, it shows the benefit to sellers from providing those services. Also, the coefficient estimates are larger than in the pooled fixed effects models. For example, in the difference in difference model for the top two sellers, the coefficient of Tmall difference is 0.4381, compared with 0.2896 in the pooled model by item. However, the reputation variables are not as intuitive. Description rating still has a positive sign and the largest coefficient. This indicates that buyers treat the description rating seriously in Chinese online markets, thus sellers providing a more accurate product description can list a higher price. Attitude rating is still statistically insignificant, meaning sellers do not think better a service attitude is an important factor in pricing strategies.

Registered sellers have specially designed pages which attract buyers. The following figures show the comparison of a fee-free seller page and a registered seller page for the same item. The item being sold here is iPhone 6 Plus. As we can see from the first figure, the registered seller has a Tmall label on the top left corner of the page. The seller's reputation information is shown on the center of the top banner. More importantly, the item has numbers of options along with advertisements on the right.

Figure 4.12 Example of list of registered seller

The fee-free seller's list has similar layout but less quality. The advertisement on the right of the page has only the pictures but no description. The options are less and without special banner on them. It is observed that successful fee-free sellers are trying to mimic the page layout of the registered sellers. However, the website keeps update the page layout and place registered sellers on top of the other sellers at majority of the time, which makes the sellers paying no additional fee hard to be found by the buyers.

Figure 4.13 Example of list of fee-free seller

Chapter 5

CONCLUSION AND DISCUSSION

Economists use search cost models to explain price dispersion as an equilibrium phenomenon. The Internet significantly reduces the search cost. Thus, price dispersion on shopping websites is predicted to disappear over time. However, the real-world data do not support this prediction. Baye, Morgan and Scholten (2004) used an information clearing house model, which is relevant when a third party provides a subset of consumers with a list of prices charged by different firms in the market. Their empirical evidence suggests that price dispersions in online markets is sizeable, pervasive, and persistent. They also found that price dispersion is negatively related to market size, using a measurement of price dispersion equal to the difference between the lowest two listing prices. Many researchers also investigated the relationship between price and the number of firms in the online market. The general finding is that high concentration is associated with higher prices. The study on eBay.com auction data showed that the seller's reputation has a positive but small impact on the price.

I use China’s online market data to explore the impact of market competition and reputation on price dispersion. Taobao.com is one of the largest online market places in the world. Using a data abstracting program, I collected and cleaned over 5.8

million listing records for 93 items in 8 categories over 3-month period from Taobao.com. The major benefit from using Taobao.com data is that recent sales volume and detailed reputation information is available. Therefore, I am able to examine the theoretical assumptions and to study the impact of market share.

The price distributions vary across different items. I take a snapshot of the price distribution for the top ranked item from each category on the first day of data collection. For some items such as the top ranked speaker, the prices are highly concentrated in a very narrow range. Though the listing price ranges from ￥10 to ￥122 for the same item, most listing prices are concentrated between ￥32.4 and ￥43.6. This is also where the top ranked seller listed its price. The result is the same when I only consider the lists which had at least one sale in last thirty days. Out of the 89 lists having at least one item sold in last month, 76 of them listed prices between ￥35.4 and ￥39.1. For some other items such as the top ranked cell phone, the prices are distributed in a relative wide range. The lowest listing price can be as low as ￥2,011, while another seller could list the same phone for ￥5,601. After removing the list without any sales in the last month, the range narrows to ￥3,000 to ￥5,505. This indicates that if the price is significantly lower than the market price, the list is very suspicious and does not get any sales.

Compared with Baye, Morgan and Scholten (2004), the data I collect from Taobao.com includes a larger number of competitors in all markets. Among the 1,000 products they observed over eight months, around 6% of their observations are single listings. Over 80% of their observations are for products advertised by 30 firms or less. Observations with more than 55 firms represent less than 0.5% of the total. In my data the number of lists for all items in any given day ranges from 55 to 17,675. The total number of lists is highly correlated with the number of lists with some sales in last month. Similarly, total sales are positively correlated with the number of lists with some sales in last month.

The Chinese online electronic markets are often shared by a group of sellers. I find that the top two sellers dominate some markets but not all. The average of the market share of the top two sellers is 53%, while it ranges from 5% in the iPhone 4S market to 93% in the Electronic Music DT-326 Headphone market. There is a negative correlation between the number of competitors and the market share of the top two sellers. The market share of the top 10 sellers is high, and it also shows a very strong negative relationship with the market size. The average market share for the top 10 sellers is 80.6%. Considering the fact that the number of sellers is typically large, this strongly suggests the markets are dominated by a small group of sellers. As the market size increases, the market share of the large sellers decreases correspondingly.

The price gap in Baye, Morgan and Scholten (2004) is not a valid measure according to real world data. First, the two lists with lowest prices do not get most of the market. I calculate the market share of the two lists with the lowest price for each item. The median of this market share is 0.03%. The highest value of this market share is 41.8% in the Edifier H180 market. However, the third highest is 5.9% indicating that the lists with the lowest two prices do not have large market share. In fact, in nearly half of the markets studied, the two lists with lowest selling prices have zero sales in last month. In fact, top sellers, in terms of sales, usually price around the medians of market prices. Among the 93 items examined, a top seller could price as low as 40% of the median market price and another top seller could price as high as 180% of the median market price. The average of the ratio between the top seller's price and the median market price is 95%. So the top sellers often price almost the same as the median market price. For 49 of the 93 items studied, the largest seller's price was higher than the second largest seller's price, while the other 44 items have the largest seller pricing lower than the second largest seller.

Second, the gap measurement in Baye, Morgan and Scholten (2004) has a very weak negative relationship with number of firms in the market. Assuming sales information is not available, the finding from Baye, Morgan and Scholten (2004) can be reproduced using China's online data. The dispersion is smaller in the larger market. However, the number of lists becomes insignificant when I control for fixed effects. The same result is found using only the lists with positive sales in the last month. This means the price gap variation within each market is small and number of lists can only explain the price gap variance between markets.

Using sales weighted coefficient of variation as the price dispersion measurement, the fixed effects price-competition regression demonstrates that the increasing number of sellers on the market leads larger dispersion. This is inconsistent with the findings documented in Baye, Morgan and Scholten (2004). Unlike the previous regressions, the coefficient is still significant even when controlling for fixed effects. The empirical result using China's online electronic data shows the opposite is true when we incorporate sales data. The sales weighted coefficient of variation measurement indicates that increasing the number of sellers in the market leads to larger dispersion. Ellison and Ellison (2009) provide a potential explanation for this. They find online sellers intend to frustrate consumer search by making the lists more complicated, causing consumers to be less price sensitive. Total number of sales negatively impacts price dispersion but the impact becomes insignificant when individual product fixed effects are included. The increasing number of competitors decreases the market concentration, as shown in Chapter 4.3.3. Dispersion-concentration models show that the decreasing market concentration, as measured by a decline in the Herfindahl index, increases the price dispersion. Holding the same level of market concentration, a change in the number of competing firms does not impact the level of price dispersion. Total sales of the market in last 30 days also has an insignificant impact on price dispersion.

I find the impact of reputation of internet sellers is small, which is similar to many other papers about US online markets. This could be caused by the fact that Chinese online sellers usually have very high ratings so the value of information about seller reputation is small. The coefficients associated with reputation are very small. My results imply that the largest increment from reputation comes from the sellers reputation for accuracy of the description of the item. A 1% increase in the description reputation relative to competitors supports a 0.0019% higher price. If a seller is the top among competitors, with a rating 100% higher than others, it will price 0.19% higher than the average price. The payback from higher reputation is negligible.

The services provided by the sellers also do not exhibit dramatic margins to the listing prices, except for the registered or authorized business sellers. It may be the case that the sellers providing the services rank higher on the search results. As a result, buyers have a higher chance to view the advertised lists and purchase from the seller who paid Taobao.com to be a registered seller. The advertising behavior has drawn attention to many researchers, such as Arnold, Li, Saliba, and Zhang (2011). Data from shopping platforms are suitable to prove the theoretical expectation or to inspire new models. However, the data for the ranking of the advertised listing information is not available in my research. I believe further studies on that topic could be very interesting.

An interesting topic for future research is to study on how sellers use bundling strategies to price discriminate between buyers. As I find from the data collection process, many lists include a gift set associated with the purchase. This is particularly true for cell phones. I find many lists, including most top sellers, offer options that bundle a phone case and other services with the phone. They are listed as options on the listing webpage so the buyers will not easily know the price for each part. The cell phone cases, as necessary accessories for expensive smart phones, can be treated as a product differentiation. A good accessory helps the seller to monopolistically compete. However, incorporating this in my analysis would require detailed information about the exact cost of the accessories. I find it very difficult to quantify the price for the accessories as they are usually not sold alone. Sellers often have private cost information about the accessories and sometimes manufacture these accessories. Perhaps in some markets the prices for accessories can be abstracted from the bundle. Then the examination of the bundling strategies will be possible.

REFERENCES

A MODEL IN WHICH AN INCREASE IN THE NUMBEROF SELLERS LEADS TO A HIGHER PRICE

Robert W. Rosenthal

ECONOMETRICA, Vol. 48, No. 6 Sep., 1980, pp. 1575-1579

A MODEL OF SALES

Hal R. Varian

THE AMERICAN ECONOMIC REVIEW, Sep. 1980

A SIMPLE MODEL OF EQUILIBRIUM PRICE DISPERSION

Jennifer F. Reinganum

JORNAL OF POLITICAL ECONOMY, Vol. 87, No. 4 Aug., 1979, pp. 851-858

ASYMMETRIC MARKET SHARES, ADVERTISING AND PRICING: EQUILIBRIUM WITH AN INFORMATION GATEKEEPER

Michael Arnold, Chenguang Li, Christine Saliba, and Lan Zhang

THE JOURNAL OF INDUSTRIAL ECONOMICS, 59(1), 2011, pp. 63-84.

AUTOMOBILE PRICES IN MARKET EQUILIBRIUM

Steven Berry, James Levinsohn and Ariel Pakes

ECONOMETRICA, Vol. 63, No. 4. Jul., 1995, pp. 841-890

BARGAINS AND RIPOFFS: A MODEL OF MONOPOLISTICALLY

COMPETITIVE PRICE DISPERSION

Steven Salop and Joseph Stiglitz

REVIEW OF ECONOMIC STUDIES, Oct. 1977, 44, pp. 493-510

BUNDLING AND COMPETITION ON THE INTERNET

Yannis Bakos and Erik Brynjolfsson

MARKETING SCIENCE, Vol. 19, No. 1, 2000, pp. 63-82

COMMODITY BUNDLING AND THE BURDEN OF MONOPOLY

William James Adams and Janet L. Yellen

THE QUARTERLY JOURNAL OF ECONOMICS, Vol. 90, No. 3, August 1976, pp.

475-498

COMPETITION AND PRICE DISPERSION IN THE US AIRLINE INDUSTRY

Severin Borenstein and Nancy L. Rose

NATIONAL BUREAU OF ECONOMIC RESEARCH, No. w3785, 1991.

COMPETITIVE PROMOTIONAL STRATEGIES

Chakravarthi Narasimhan

THE JOURNAL OF BUSINESS, Vol. 61, No. 4 Oct., 1988, pp. 427-449

CONSUMER DECISION-MAKING AT AN INTERNET SHOPBOT: BRAND STILL MATTERS

Michael D. Smith and Erik Brynjolfsson

THE JOURNAL OF INDUSTRIAL ECONOMICS, 49 Dec. 2001, pp. 541-558

DOES A SELLER'S ECOMERCE REPUTATION MATTER? EVIDENCE FROM EBAY AUCTIONS

Mikhail I. Melnik and James Alm

THE JOURNAL OF INDUSTRIAL ECONOMICS, Vol. 50, No. 3 Sep. 2002

DOES COMPETITION REDUCE PRICE DISPERSION? NEW EVIDENCE FROM THE AIRLINE INDUSTRY

K. S. Gerardi and A. H. Shapiro

JOURNAL OF POLITICAL ECONOMY, Vol 117(1), 2009, pp. 1-37

DOES THE INTERNET MAKE MARKETS MORE COMPETITIVE? EVIDENCE FROM THE LIFE INSURANCE INDUSTRY

Jeffrey R. Brown and Austan Goolsbee

JOURNAL OF POLITICAL ECONOMICS. 2002, vol. 110. no.3

ENDOGENEITY IN THE CONCENTRATION-PRICE RELATIONSHIP: CAUSES, SONSEQUENCES AND CURES

William N. Evans, Luke M. Froeb and Gregory J. Werden

THE JOURNAL OF INDUSTRIAL ECONOMICS, 1993, 61, 431-438

EQUILIBRIUM COMPARISON SHOPPING

Louis L. Wilde and Alan Schwartz

REVIEW OF ECONOMIC STUDIES, July 1979, 46, pp. 543-554

HORIZONTAL MERGERS OF ONLINE FIRMS: STRUCTURAL ESTIMATION AND COMPETITIVE EFFECTS

Yonghong An, Michael R. Baye, Yingyao Hu, Matt Shum and John Morgan

MIMEO

INFORMATION AND MONOPOLISTIC COMPETITION

S. Salop

THE AMERICAN ECONOMIC REVIEW, Vol. 66, No. 2, Papers and Proceedings of the Eighty-eighth Annual Meeting of the American Economic Association (May, 1976), pp. 240- 245

INFORMATION, SEARCH, AND PRICE DISPERSION

Michael R. Baye and John Morgan

HANDBOOK OF ECONOMICS AND INFORMATION SYSTEMS, 2006, Volume 1

MODELS OF MARKET ORGANIZATION WITH IMPERFECT INFORMATION: A SURVEY

Michael Rothschild

JOURNAL OF POLITICAL ECONOMY, Vol. 81, No. 6 Nov. - Dec., 1973, pp. 1283-1308

NUMBER OF SELLERS, AVERAGE PRICES, AND PRICE DISPERSION

J. M. Barron, B.A. Taylor and J. R. Umbeck

INTERNATIONAL JOURNAL OF INDUSTRIAL ORGANIZATION, Vol 22(8), 2004, pp. 1041-1066.

PENNIES FROMEBAY: THE DETERMINANTS OF PRICE IN ONLINE AUCTIONS

David Lucking-Reiley, Doug Bryan, Naghi Prasad and Daniel Reeves

THE JOURNAL OF INDUSTRIAL ECONOMICS, Volume LV June 2007 No. 2

PRICE AND PRICE DISPERSION ON THE WEB: EVIDENCE FROM THE ONLINE BOOK INDUSTRY OF TAIWAN

Yu-Chen Lin and Chiang-Ming Chen

HITOTSUBASHI JOURNAL OF ECONOMICS, Vol. 55, No. 1 June 2014, pp. 51-70

PRICE CONCENTRATION STUDIES: THERE YOU GO AGAIN

C. Newmark

“CONCENTRATION AND MARKET SHARES” PANEL, 2004.

PRICE DISPERSION AND COMPETITION WITH DIFFERENTIATED SELLERS

Matthew Lewis

THE JOURNAL OF INDUSTRIAL ECONOMICS, Vol 56, no. 3 2008, pp. 654-678

PRICING AND MARKET CONCENTRATION IN OLIGOPOLY MARKETS

Vishal Singh and Ting Zhu

MARKETING SCINCE, Vol. 27, No. 6, November–December 2008, pp. 1020–1035

PRICE AND PRICE DISPERSION ON WEB: EVIDENCE FROM THE ONLINE BOOK INDUSTRY

Karen Clay, Ramayya Krishnan and Eric Wolff

THE JOURNAL OF INDUSTRIAL ECONOMICS, 49, Dec. 2001, pp. 521-539

PRICE DISPERSION IN THE SMALL AND IN THE LARGE: EVIDENCE FROM AN INTERNET PRICE COMPARISON SITE

Michael R. Baye, John Morgan and Patrick Scholten

THE JOURNAL OF INDUSTRIAL ECONOMICS, Volume LII December 2004 No. 4

QUALITY EXPECTATIONS, REPUATION AND PRICE

Stuart Landon and Constance E. Smith

SOUTHERN ECONOMIC JOURNAL, Vol. 64, No. 3, Jan., 1998, pp. 628-647

REPUTATION AND PRICES ON THE E-MARKET: EVIDENCE FROM A MAJOR FRENCH PLATFORM

G. Jolivet, B. Jullien and F. Postel-Vinay

INTERNATIONAL JOURNAL OF INDUSTRIAL ORGANIZATION, Vol 45, 2016, pp. 59-75.

REPUTATION IN AUCTIONS: THEORY, AND EVIDENCE FROM EBAY

D. Houser and J. Wooders

JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY, 15(2), 2006, pp. 353-369

SEARCH AND MARKET EQUILIBRIUM

Richard D. MacMinn

JOURNAL OF POLITICAL ECONOMY, Vol. 88, No. 2 Apr., 1980, pp. 308-327

SEARCH, OBFUSCATION, AND PRICE ELASTICITIES ON THE INTERNET

Glenn Ellison and Sara F. Ellison

ECONOMETRICA, Vol. 77, No 2, 2009, pp. 427-452

THE ECONOMICS OF INFORMATION

George J. Stigler

JOURNAL OF POLITICAL ECONOMY, Vol. 69, No. 3 (Jun., 1961), pp. 213-225

THE MARKET OF 'LEMONS': QUALITY UNCERTAINTY AND MARKET MECHANISM

George Akerlof

THE QUARTERLY JOURNAL OF ECONOMICS, 1970, 84, pp. 488-500

THE THEORY OF SALES: A SIMPLE MODEL OF EQUILIBRIUM PRICE DISPERSION WITH IDENTICAL AGENTS

S. Salop and J. E. Stiglitz

THE AMERICAN ECONOMIC REVIEW, Vol. 72, No. 5 (Dec., 1982), pp. 1121-1130

THE ROLE OF MARKET FORCES IN ASSURING CONTRACTUAL PERFORMANCE

Benjamin Klein and Keith B. Leffler

THE JOURNAL OF POLITICAL ECONOMY, 1981, pp. 615-641.

THE WINNER’S CURSE, RESERVE PRICES AND ENDOGENOUS ENTRY: EMPIRICAL INSIGHTS FROM EBAY AUCTIONS

Patrick Bajari and Ali Hortaçsu

THE RAND JOURNAL OF ECONOMICS, Vol. 34, No. 2 Summer, 2003, pp. 329-355

TRUST AMONG STRANGERS IN INTERNET TRANSACTIONS: EMPIRICAL ANALYSIS OF EBAY’S REPUTATION SYSTEM

P. Resnick and R. Zeckhauser

THE ECONOMICS OF THE INTERNET E-COMMERCE, Vol 11(2), 2002, pp. 23-25.

WHY BUNDLE DISCOUNTS CAN BE A PROFITABLE ALTERNATIVE TO COMPETING ON PRICE PROMOTIONS

Subramanian Balachander, Bikram Ghosh and Axel Stock

MARKETING SCIENCE, Vol. 29, No. 4, July-August 2010, pp. 624-638

Appendix A

T TEST FOR PRICES OF LISTS BY SELLER WITH SALES VERSUS THOSE WITHOUT (CONT.)

Table A.1: t test for prices of lists by seller with sales versus those without

Sennheiser MX80	Headphone	64.0	75.0	-42.6	0.00
Somic ST-2688	Headphone	33.8	39.1	-13.0	0.00
Sony MDR-EX10A	Headphone	58.2	73.7	-33.6	0.00
Acer E1-471G-53212G50Mn	Laptop	3,002.6	3,444.8	-49.5	0.00
Asus A45EI323VD-SL	Laptop	3,589.2	3,810.7	-44.4	0.00
Asus X401EI235A	Laptop	2,737.9	3,003.5	-34.0	0.00
Asus X45EI237VD-SL	Laptop	2,682.0	2,873.8	-56.0	0.00
Dell Ins15r-978	Laptop	2,241.4	2,703.8	-15.2	0.00
Hasee K580S-I7 D0	Laptop	4,141.7	4,497.2	-28.6	0.00
HP dv6-6029TX	Laptop	3,180.3	3,345.3	-7.5	0.00
Lenovo G480-IFI	Laptop	3,244.8	3,622.0	-35.4	0.00
Lenovo Y400N-IFI	Laptop	4,904.0	5,065.2	-33.7	0.00
ThinkPad E430c(33651B8)	Laptop	3,516.2	3,696.3	-9.4	0.00
ACER DS-1005	Mouse	32.3	38.1	-10.8	0.00
Cherry JM-0300	Mouse	102.5	101.0	2.1	0.03
Dell MS111	Mouse	30.0	35.7	-48.8	0.00
Lenovo M20	Mouse	22.4	27.7	-22.9	0.00
Lenovo ThinkPad 57Y4635 black mouse	Mouse	47.1	53.5	-32.6	0.00
Logitech G1	Mouse	57.0	101.7	-81.9	0.00
Logitech G500	Mouse	314.8	345.6	-32.2	0.00
Logitech M100	Mouse	52.3	59.7	-46.7	0.00
Pennefather N6000	Mouse	30.1	37.3	-48.6	0.00
Razer / Razer DeathAdder upgraded version	Mouse	225.6	255.0	-34.9	0.00
Razer / Razer hell mad snake mirror version	Mouse	112.7	148.5	-50.8	0.00
Apple iPhone 4S	Phone	3,179.2	3,747.1	-161.5	0.00
Apple iPhone 5	Phone	4,250.5	4,662.4	-78.7	0.00
Daxian GS5000	Phone	93.8	122.9	-44.0	0.00
Daxian W111	Phone	85.5	119.7	-66.7	0.00
Huawei G520	Phone	620.7	882.3	-223.8	0.00
Lenovo A820T	Phone	613.8	813.8	-155.1	0.00
MIUI 2A(MI2A)	Phone	1,532.9	1,614.9	-49.6	0.00
MIUI 2S(MI2S)	Phone	1,881.6	2,051.3	-94.2	0.00
Nokia 1050	Phone	144.0	185.2	-62.3	0.00
Nokia 1120	Phone	122.0	201.6	-98.4	0.00
Nokia 2030	Phone	78.0	112.3	-60.6	0.00
Samsung GALAXY S4 I9500	Phone	3,606.8	4,077.6	-163.5	0.00
Samsung I9300 GALAXY SIII	Phone	2,430.2	2,934.6	-206.1	0.00
Cruiser R101T06	Speaker	147.0	157.6	-38.8	0.00
Cruiser R10U	Speaker	70.1	78.5	-18.2	0.00
Cruiser R201T08	Speaker	197.3	204.3	-19.6	0.00
Dell AX210 USB	Speaker	52.4	61.4	-24.0	0.00
Dell AX510	Speaker	97.8	136.3	-46.6	0.00
Edifier R101V	Speaker	128.9	134.8	-14.6	0.00
Edifier R151T	Speaker	306.2	319.8	-19.5	0.00
JBL Duet	Speaker	121.1	169.8	-15.7	0.00
Lenovo Lenovo L1520	Speaker	51.3	116.9	-97.6	0.00
Microlab M-200 tenth anniversary edition	Speaker	247.7	264.7	-25.3	0.00
Philips SPA1312	Speaker	135.3	145.4	-15.2	0.00
Sound Pai think S020	Speaker	45.2	67.3	-40.7	0.00
Apple iPad mini(16G)WIFI	Tablet	2,216.5	2,366.5	-40.8	0.00
Apple iPad mini(32G)WIFI	Tablet	2,848.7	2,973.9	-17.8	0.00
Apple iPad4(16G)WIFI	Tablet	3,215.4	3,324.3	-30.3	0.00
Apple iPad4(32G)WIFI	Tablet	3,767.5	3,915.1	-25.7	0.00
CUBE U25GT (8G) WIFI version	Tablet	314.4	366.6	-24.1	0.00
Lenovo Ideatab A1000(4G)	Tablet	801.3	836.1	-17.1	0.00
Lenovo Pad A1(16G)	Tablet	855.4	1,000.6	-5.3	0.00
Samsung galaxy note 10.1N8000(16G)	Tablet	2,479.8	3,378.3	-57.8	0.00
Samsung Galaxy Tab P3100(8G)3G	Tablet	1,461.9	1,883.8	-48.9	0.00
Taipower P85 dual-core(16G)	Tablet	518.5	630.6	-34.8	0.00
Taipower P88 quad-core (16G)	Tablet	802.1	854.2	-14.1	0.00
window N70S (8G)WIFI	Tablet	375.4	453.5	-24.3	0.00

Appendix B

NUMBER OF LISTINGS BY PRICE FOR SELECTED ITEM

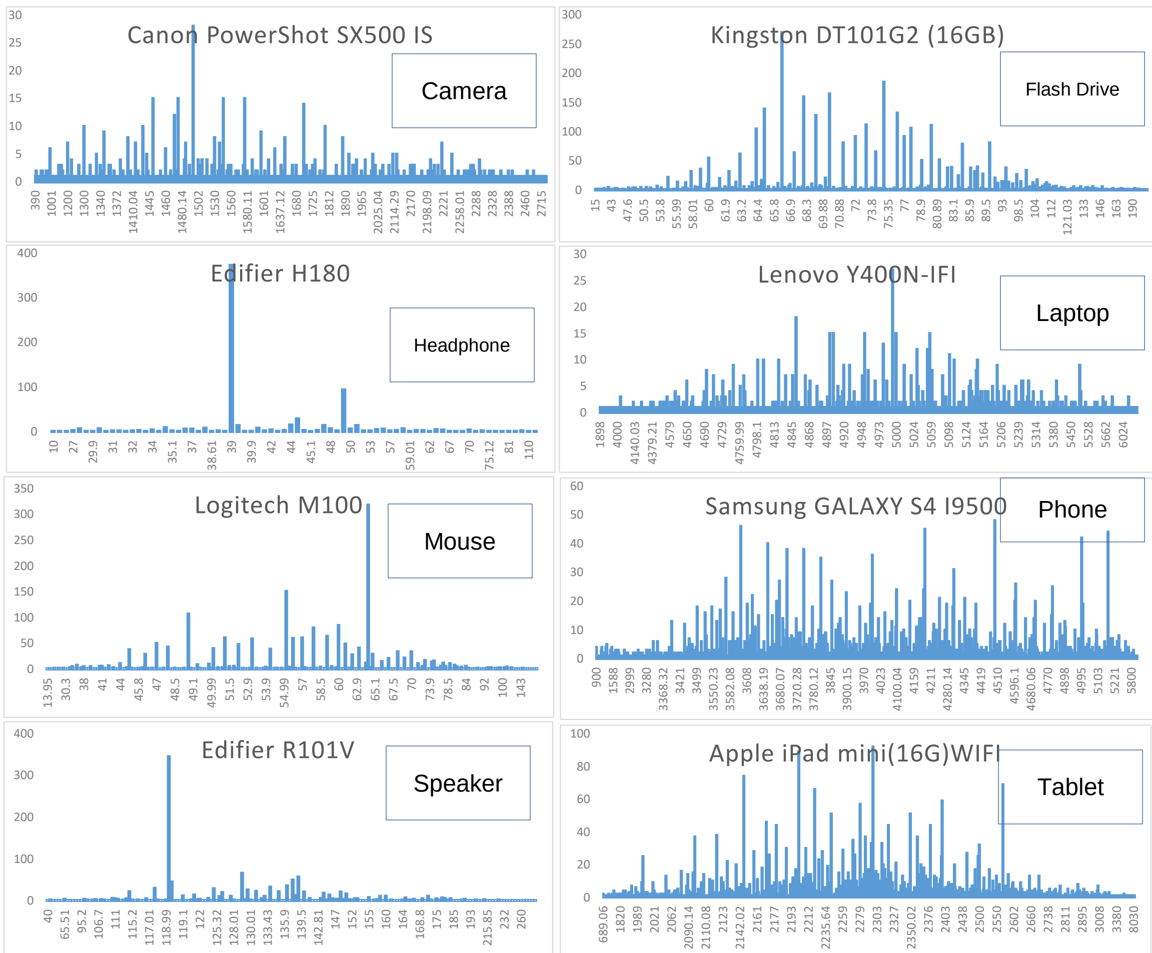


Figure B.1 Number of listings by price for selected item

Appendix C

NUMBER OF LISTINGS BY PRICE FOR SELECTED ITEM, TAKING OUT LISTINGS WITH ZERO SALES IN LAST THIRTY DAYS

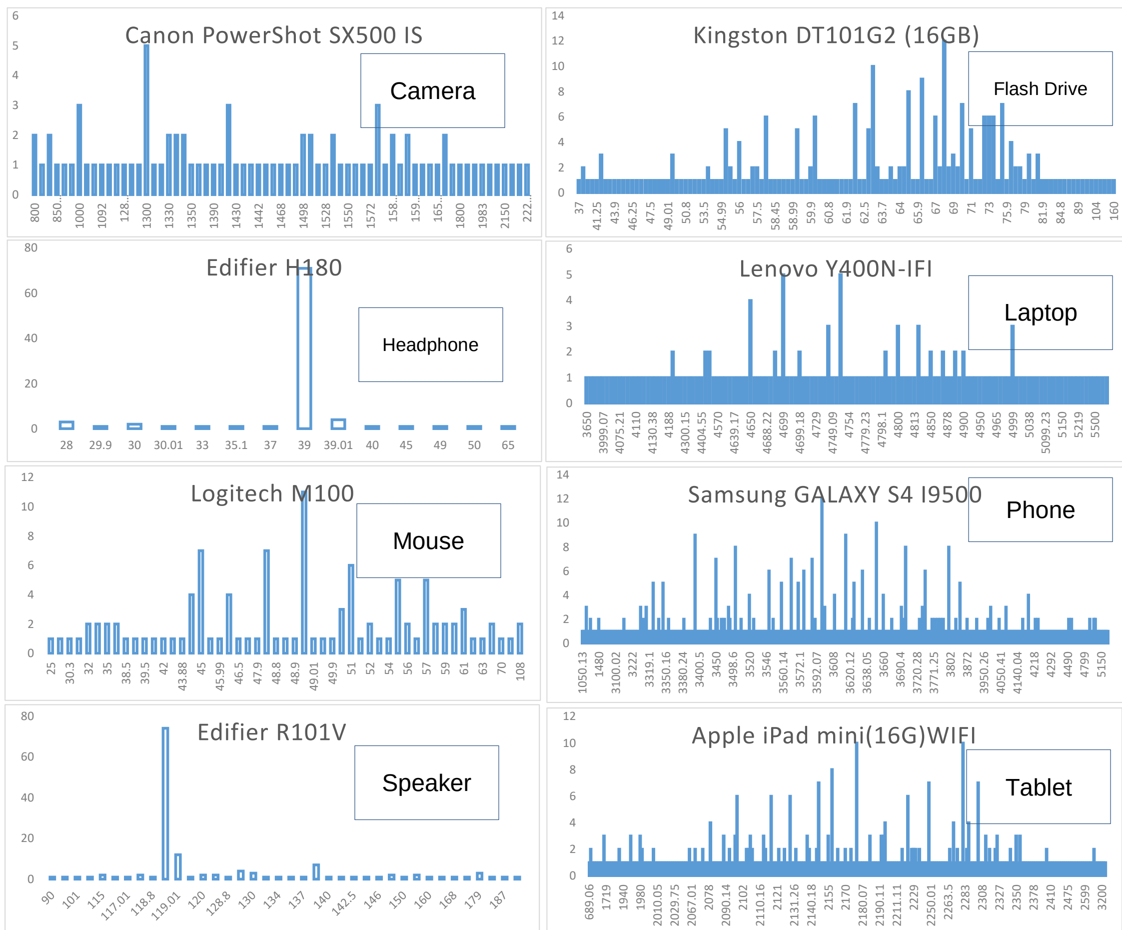


Figure C.1 Number of listings by price for selected item, taking out listings with zero sales in last thirty days