Proceedings of the 2nd International Conference on Business and Policy Studies
DOI: 10.54254/2754-1169/8/20230335

A Measurement of Price Dispersion and Search Cost in China's Online Stores

Zian Liu[1,a], Zewen Wang[2,b,*]

[1]Harrow Beijing, Beijing,100102, China
[2]United World College Changshu China, Changshu, 215500, China
a. zianliu@harrowbeijing.cn, b. 3252216130@qq.com
**corresponding author*

Abstract: We studied a measurement of price dispersion and search cost to find implications of relative search cost. We investigated several models, including the classical search cost model, price dispersion estimation model and Stahl's consumers' gain model. According to the data we collected, the empirical analysis explores absolute search cost, relative search cost, consumers' gain from search, competitive prices, and comment-weighted gradings based on the settings. With the hypothesis that we clarified and statistics applications, this paper attempts to illustrate that price dispersion among online book retailers increases as price rises. Besides, we testified to the effects of information and discovered that online shopping isn't frictionless. Last, we tested the correlation between gradings and search cost to testify our conclusions. The results are statistically significant under a 5% p-value.

Keywords: price dispersion, search cost, Stahl's model, reservation price

1. Introduction

Search cost for consumers is defined as the cost of gaining information in a market [1]. Due to imperfect understanding of actual supply and demand conditions, price dispersion arises. If the search cost of a good is absolutely fixed regardless of its price, consumers should spend similar input in searching. This statement is contradictory to our daily observation. For example, we generally do more search on computers than draft papers. However, if the search cost is relative to price, it's important to testify their relationship and to discuss the further economic inspirations.

2. Literature Review

In mainstream literature, there are different and related findings about consumer search. "Testing models of consumer search using data on web browsing and purchasing behaviour" confirms that fixed sample size search provides a more realistic description of search behaviour than the benchmark model of sequential search with a known price distribution based on recall patterns observed in the data and the absence of dependence of search decisions on prices by using a big dataset of the online shopper and consumer behaviour [2].

Besides, "Oligopolistic pricing with sequential consumer search." by Stahl states that while several identical stores compete, customers search consecutively and with flawless recollection for the best deal. All other consumers have positive search costs, except for shoppers, who have zero

DOI: 10.54254/2754-1169/8/20230335

search costs. A unique symmetric Nash equilibrium price distribution is found with certain properties. “As the number of stores increases, the Nash equilibrium becomes more monopolistic” [3].

Moreover, "Big data: New tricks for econometrics." in the Journal of Economics discovered that big data would continue to expand, and small dataset data modification technologies are insufficient to handle new issues [4]. In addition, the paper “Information gatekeepers and price discrimination on the internet.” by Baye, Michael R., and John Morgan “examine the equilibrium interaction between a market for price information and the homogeneous product market it serves” [5]. Gatekeeper will maximize its profits in an equilibrium satisfying one of four certain conditions related to price dispersion, access fee, advertising fees and prices.

In short, these main papers strengthen our understanding of such topics, inspire us to develop our thoughts and gain our findings.

3. Models

Absolute search cost model:

$$Search\ cost = r - c$$

Stahl's consumers' gain model:

$$Gain = \frac{r + s}{2}$$

Price dispersion estimation model:

$$Search\ cost = \frac{r + c}{p}$$

Variables are s (search cost) and r (reservation price)

Where variable p stands for market price level.

3.1. Data Collection

We focused on the online book retailers since the sellers have fewer concerns with production costs and more power to determine their products. We observed the purchasing data from DangDang, one of the largest Chinese online book-selling platforms. For each product, the independent variables are price, number of comments, and gradings. We recorded each variable of a homogenous book from five different retailers. Eventually, we obtained 6,000 data from 300 books to reduce uncertainty and random errors.

3.2. Variables

An important assumption is that the book retailers are in imperfect competition, in which each seller has a certain level of market power of a homogenous product. Moreover, the distribution of equilibrium price is assumed to be uniformly distributed to simplify the further calculation.

Moreover, it is assumed that there isn't such a consumer that searches costlessly or doesn't search at all. Thus, we can ensure that the reservation price is always higher than the competitive price, and consumers have incentives to search for a better price with certain costs.

The two most important variables are reservation price and competitive price. Since we can only observe the information that is accessible to the consumers, the maximum price is set to be the reservation price, and the minimum price is set to be the competitive price.

For each product, the market price is equal to the average price. After substitution and

DOI: 10.54254/2754-1169/8/20230335

manipulation, here are the empirical models:

Absolute search cost model:

$$Search\ cost = P_{max} - P_{min}$$

Stahl's consumers' gain model:

$$Gain = \frac{P_{max} + P_{min}}{2}$$

Price dispersion estimation model:

$$Search\ cost = \frac{P_{max} + P_{min}}{P_{avg}}$$

P stands for prices.

3.3. Absolute Search Cost

Absolute search cost exists as a fixed constant of a product regardless of its price. We observed that books have positive search cost, as represented by the y-intercept in Figure 1. This suggests that the absolute search cost of a product is valid when the price changes incrementally. However, the search cost of a product might change when the price is magnified by a scale factor (Table 1).

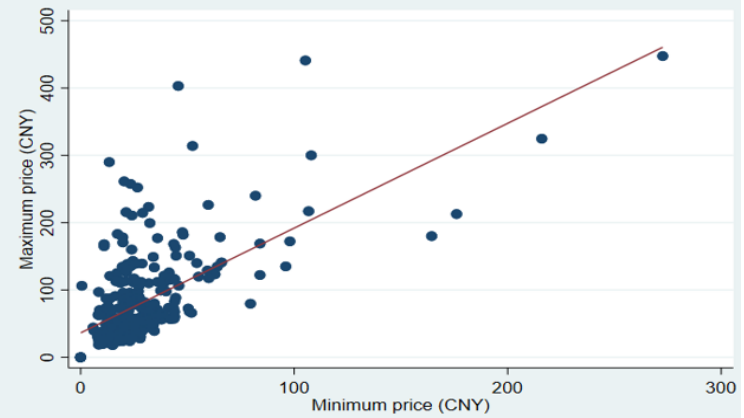

Figure 1: Regression model graph of absolute search cost.

Table 1: Regression model data of absolute search cost.

Maximumprice	Coefficient	Std. err.	t	$p > \|t\|$	$[95\%\ conf.interval]$	
Minimumprice	1.556882	.1108864	14.04	0.000	1.338662	1.775101
_cons	36.20347	4.390237	8.25	0.000	27.56367	44.84326

For example, when the maximum price and the minimum price are magnified by 10 times, the absolute search cost is also magnified by the same amount in Table 2. This indicates that absolute search cost might be inaccurate when the prices are changing by scale.

DOI: 10.54254/2754-1169/8/20230335

Table 2: Regression model of absolute search cost after 100x magnification.

Maximumprice10	Coefficient	Std. err.	t	$p > \|t\|$	[95% $conf.interval$]	
Minimumprice10	1.556882	.1108864	14.04	0.000	1.338662	1.775101
_cons	362.0347	43.90237	8.25	0.000	275.6367	448.4326

3.4. Relative Search Cost

Noticing the constraints of absolute search cost, we hypothesized that search cost is relative to prices. We used price dispersion to estimate the search cost of each product. Search cost is the main factor that causes price dispersion as it gives firms monopoly power.

We found that price dispersion increases as the market price of a book increases in Figure 2. This indicates that the search cost of a product will increase comparatively along with its price (Table 3 and Table 4).

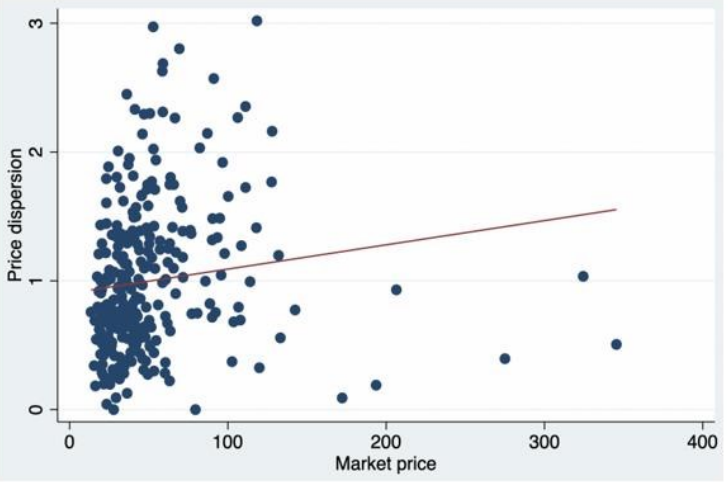

Figure 2: Regression model price dispersion estimation graph.

Table 3: Regression model price dispersion estimation data.

ProxyFrict~x	Coefficient	Std. err.	t	$p > \|t\|$	[95% $conf.interval$]	
Averageprice	.0018816	.0008391	2.24	0.026	.0002302	.0035331
_cons	.9033605	.0539795	16.74	0.000	.7971208	1.0096

The main contributor to this correlation is the different elasticities of the reservation price (maximum price) and the competitive price (minimum price). The reservation price of a product is more responsive than the competitive price when its price changes in Figure 3. Changes in the maximum prices are more sensitive to the changes in the actual market prices. This makes the price dispersion increase as the market price increases. As a result, the estimation is that relative search cost is positively correlated to the average market price.

DOI: 10.54254/2754-1169/8/20230335

This result potentially implies that the price of a product itself could somehow provide consumers with extra information that adds to its value. The most relative attribute that the price of a product implies to the consumers is its quality. Higher prices often give consumers impressions of higher quality, which can change consumers' demand and affect the minimum price and the maximum price. When a product is considered to have higher quality, the consumers will believe that the demand for the product has increased. Thus, they will boost the actual demand as most of the consumers are more willing to purchase the product at higher price spontaneously.

Higher demand due to the implication from prices will increase the reservation price at each price level. At the same time, the competitive prices will decrease because greater demand incentivises more buyers and sellers to enter the market. Moreover, incentivised searching will also press the competitive price lower.

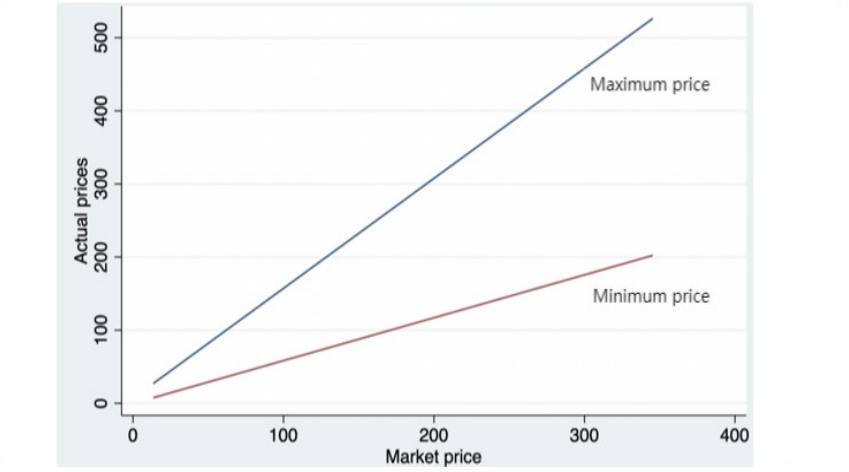

Figure 3: Regression model graph of maximum price and minimum price.

Table 4: Regression model data of maximum price and minimum price.

Minimumprice	Coefficient	Std. err.	t	$p > \|t\|$	$[95\% \; conf. interval]$	
Averageprice	.5865056	.0207646	28.25	0.000	.5456377	.6273735
_cons	-.311607	1.335836	-0.23	0.816	-2.940732	2.317518
Maximumprice	Coefficient	Std. err.	t	$p > \|t\|$	$[95\% \; conf. interval]$	
Averageprice	1.504391	.044307	33.95	0.000	1.417188	1.591594
_cons	6.742746	2.850372	2.37	0.019	1.132787	12.35271

4. Consumers' Gain from Search

Consumers' gain from search varies with search cost. We hypothesized that consumers gain more benefits under higher search costs. Following economic intuition, consumers' gain from search is highly dependent on the difference between the reservation price and the actual price. In this case, consumers gain benefit from search because they can find better prices that are better than the previous ones. We used Stahl's model to determine consumers' gain from search empirically (Table 5).

DOI: 10.54254/2754-1169/8/20230335

$$Gain = \frac{1}{r-c}\int_c^r f(x) = (\frac{r^2}{2} - \frac{c^2}{2})(r - \frac{1}{c})$$

On the aim of simplifying data analysis, we assumed that the equilibrium price is uniformly distributed in integration. Thus, we have:

$$Gain = \frac{r+c}{2}$$

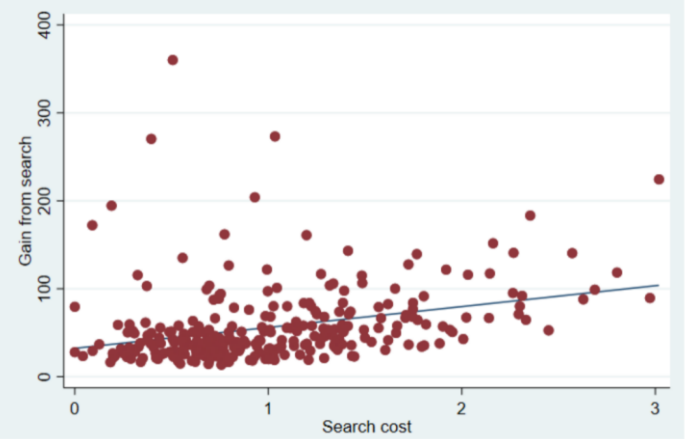

Figure 4: Regression model graph of consumers' gain from search.

Table 5: Regression model data of consumers' gain from search.

Gain	Coefficient	Std. err.	t	$p > \|t\|$	[95% $conf.interval$]	
ProxyFrictionMax	23.72944	4.192638	5.66	0.000	15.4777	31.98118
_cons	32.28805	4.832208	6.68	0.000	22.77754	41.79856

We found that there is a strong positive correlation between consumers' gain from search and search friction in Figure 4. The cause of this correlation is the simultaneous increase in both competitive prices and reservation prices. However, increasing consumers' gain from search could mean a greater opportunity cost of not searching in other words. Thus, as relative search cost increases, the inequality between shoppers and "non-shoppers" is exacerbated.

5. Robustness Check

Besides prices, other accessible information about the product may also contribute to the formation of search costs. We used the number of comments on each product to check the validity of the assumption of relative search cost. Since each comment is matched to grading, the grading for each product is weighted holistically.

Thus, here is the econometric model that we constructed to test the relationship between gradings and relative search cost:

$$\text{Search cost} = (\frac{Individual\ comment}{Total\ comment}\text{x grading})\text{Xi} + \text{E}$$

The hypothesis made for the gradings is that it is positively correlated to price dispersion, as well

DOI: 10.54254/2754-1169/8/20230335

as relative search cost. However, different gradings could affect the shoppers differently. We assumed that positive comments can encourage consumers to purchase, and vice versa. When consumers are willing to purchase a product with higher prices, they have fewer incentives to search for a better price because the current price is relatively "better". Although the difference between the reservation price and the current price is larger, positive comments also increase search costs by discouraging searching. Thus, positive comments could impede consumers' search without being aware because consumers won't feel bad for buying a highly rated product at a higher price.

Out of 5, we classified grades 2.5 - 5 as positive grades and 0 - 2.5 as negative. Since the products are generally highly rated in Table 6, we are focused to testify whether higher grades could cause larger price dispersion. Thus, we also expect that higher grades could lead to higher search costs.

Moreover, these gradings are being weighted due to the platform mechanism in that each grading is released with a comment. Since the influence of grades is mainly dependent on their number on each product, we weighted each comment to the total number to ensure that the more comments, the more influence on consumers.

Table 6: Summary statistics of gradings.

Variable	Obs	Mean	Std. dev.	Min	Max
GradingAct~1	245	4.583211	.7482143	0	5

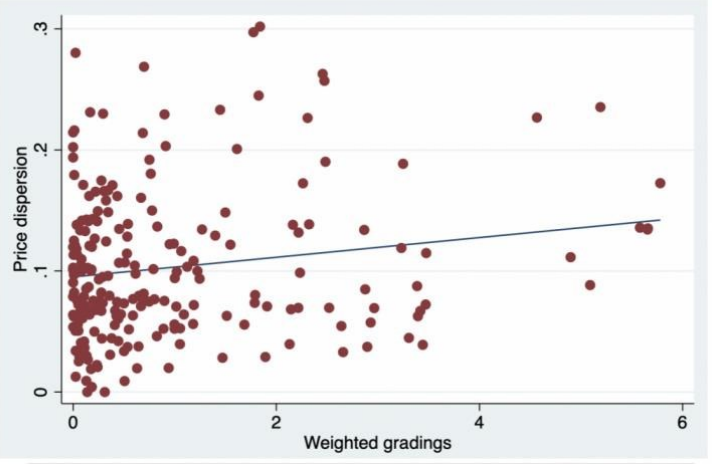

Figure 5: Regression model graph of price dispersion and gradings.

Table 7: Regression model data of price dispersion and gradings.

MaxF10	Coefficient	Std. err.	t	$p > \|t\|$	[95% $conf.interval$]	
WeightedGrading	.0081394	.0032828	2.48	0.014	.0016689	.0146098
_cons	.0950716	.0051687	18.39	0.000	.084884	.1052591

DOI: 10.54254/2754-1169/8/20230335

Throughout statistical analysis, we found that more positive comments can lead to higher price dispersion in Figure 5, as well as relative search cost. This result proves that our previous hypothesis of gradings is valid and supports our estimation of the relative search cost of books (Table 7).

The main contributor to the positive correlation between search cost and grades is the decreasing competitive price when search friction increases in Table 8. It's plausible to say that more positive information can boost the demand for a book, which increases price elasticity and lowers the competitive price.

Table 8: Dynamics of minimum price and maximum price.

ProxyFrict~x	Coefficient	Std. err.	t	$p > \|t\|$	[95% $conf.interval$]	
Minimumprice	-.0047001	.001204	-3.90	0.000	-.0070698	-.0023304
_cons	1.135982	.0482364	23.55	0.000	1.041046	1.230919
ProxyFrict~x	Coefficient	Std. err.	t	$p > \|t\|$	[95% $conf.interval$]	
Maximumprice	.0042681	.000436	9.79	0.000	.00341	.0051262
_cons	.6454917	.0464378	13.90	0.000	.5540951	.7368883

On the other hand, there's an opposing force that reduces relative search costs by increasing reservation prices. When the reservation price increases, the competitive price is less effective in the model of relative search cost as being a more diluted numerator. However, this effect is smaller than the positive effect on relative search cost, in which the absolute value of the coefficient of the minimum price is larger. This result suggests that the reservation price is comparatively inelastic than the competitive price.

6. Conclusion

Our observation of China's online book retailers provides an implication of the existence of relative search cost, which is different to the absolute search cost. Thus, we proved that the search cost of a homogenous is correlated to its price. The results indicate that the reservation price and the competitive price of the consumers are dynamic when the market price level changes. This finding implies that the price of a product could change its search cost by providing consumers with supplementary information, which opposes the idea that search cost and price are unrelated.

Moreover, the results of price dispersion from statistical analysis prove Stahl's relative search cost empirically under several simplifying assumptions. Since consumers can gain more benefits from searching as the price dispersion gets larger, it's easier to find a lower price when the relative search cost increases. At the same time, searching can magnify the inequality between the costless shoppers and the costly shoppers. The maximum difference between them is represented by the difference between the reservation price and the competitive price.

Additionally, we proved again that online shopping isn't frictionless as advertised on the social media. Also, it's more necessary to search for lower prices when the price of a product increases.

7. Limitations

First, we can only observe the data that are accessible to the consumers, which prevents us from analyzing the actual behaviour patterns (e.g., number of clicks and number of browses). Therefore,

DOI: 10.54254/2754-1169/8/20230335

our analysis is influenced by the marketing strategies of the platforms. For example, we only observed the books with positive gradings because the undesirable books are either hidden or out of stock. Also, it's generally observed on all platforms that the sellers would improve their impressions to the buyers by faking positive grades and comments.

Second, the conclusions might not be applicable in the other markets. Different to books, the prices of other products are more restricted by the markup of its production costs, which reduces price dispersion as well as the relative search cost. Also, the prices for books are generally cheap and within a small interval. This increases the uncertainty in our measurement of prices and makes the results unreliable in the markets with higher prices (e.g., machinery, electronic devices, and luxury).

Third, we only used a linear regression model in the econometric analysis. The relationship between variables is more complicated than affecting a single gradient. Consequently, the correlations should be weaker than the actual values.

Due to the previous limitations, our derivation of relative search cost is only an inaccurate reflection of price dispersion since the maximum price and the minimum price can't represent the reservation price and the competitive price accurately. Moreover, we assumed that the equilibrium price in Stahl's model is uniformly distributed, which deviates from the real-life situation while simplifying the process of data analysis.

Acknowledgement

Zian Liu and Zewen Wang contributed equally to this work and should be considered co-first authors.

References

[1] *Stigler, G. J. (1961). The Economics of Information. Journal of Political Economy, 69(3), 213–225. https://doi.org/10.1086/258464.*

[2] *De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest. "Testing models of consumer search using data on web browsing and purchasing behavior." American economic review 102.6 (2012): 2955-80.*

[3] *Stahl, Dale O. "Oligopolistic pricing with sequential consumer search." The American Economic Review (1989): 700-712.*

[4] *Varian, Hal R. "Big data: New tricks for econometrics." Journal of Economic Perspectives 28.2 (2014): 3-28.*

[5] *Baye, Michael R., and John Morgan. "Information gatekeepers and price discrimination on the internet." Economics Letters 76.1 (2002): 47-51.*