

*JSPS Grants-in-Aid for Scientific Research (S)*  
*Understanding Persistent Deflation in Japan*

*Working Paper Series*

*No. 007*

*April 17, 2013*

---

Why are product prices in online markets not converging?

Takayuki Mizuno  
Tsutomu Watanabe

---

UTokyo Price Project  
702 Faculty of Economics, The University of Tokyo,  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: +81-3-5841-5595  
E-mail: [watlab@e.u-tokyo.ac.jp](mailto:watlab@e.u-tokyo.ac.jp)  
<http://www.price.e.u-tokyo.ac.jp/english/>

Working Papers are a series of manuscripts in their draft form that are shared for discussion and comment purposes only. They are not intended for circulation or distribution, except as indicated by the author. For that reason, Working Papers may not be reproduced or distributed without the expressed consent of the author.

# Why are product prices in online markets not converging?

Takayuki Mizuno<sup>1,2,3,\*</sup>, Tsutomu Watanabe<sup>2,3</sup>

1 National Institute of Informatics, Tokyo, Japan

2 Graduate School of Economics, University of Tokyo, Tokyo, Japan

3 The Canon Institute for Global Studies, Tokyo, Japan

\* E-mail: mizuno@nii.ac.jp

## Abstract

Why are product prices in online markets dispersed in spite of very small search costs? To address this question, we construct a unique dataset from a Japanese price comparison site, which records price quotes offered by e-retailers as well as customers' clicks on products, which occur when they proceed to purchase the product. We find that the distribution of prices retailers quote for a particular product at a particular point in time (divided by the lowest price) follows an exponential distribution, showing the presence of substantial price dispersion. For example, 20 percent of all retailers quote prices that are more than 50 percent higher than the lowest price. Next, comparing the probability that customers click on a retailer with a particular rank and the probability that retailers post prices at a particular rank, we show that both decline exponentially with price rank and that the exponents associated with the probabilities are quite close. This suggests that the reason why some retailers set prices at a level substantially higher than the lowest price is that they know that some customers will choose them even at that high price. Based on these findings, we hypothesize that price dispersion in online markets stems from heterogeneity in customers' preferences over retailers; that is, customers choose a set of candidate retailers based on their preferences, which are heterogeneous across customers, and then pick a particular retailer among the candidates based on the price ranking.

## 1 Introduction

The number of internet users worldwide is 2.4 billion, constituting about 35 percent of the global population. The number of users has more than doubled over the last five years and continues to increase [1]. In the early stages of the internet boom, observers predicted that the spread of the internet would lead the retail industry toward a state of perfect competition, or a Bertrand equilibrium [2]. For instance, *The Economist* stated in 1990 that “[t]he explosive growth of the Internet promises a new age of perfectly competitive markets. With perfect information about prices and products at their fingertips, consumers can quickly and easily find the best deals. In this brave new world, retailers’ profit margins will be competed away, as they are all forced to price at cost” [3]. Even academic researchers argued that online markets will soon be close to perfectly competitive markets [4][5][6][7].

Has this prediction come true? Unfortunately not. Even now, e-retailers quote different prices for a particular product, and those who quote prices above the lowest price still survive in the market. This is reflected in empirical studies on a variety of products showing that a wide dispersion in the prices quoted by e-retailers can be observed [7][8][9][10][11][12]. An important

implication of the existence of such a wide price dispersion is that customers do not make their purchase decisions on the basis of product prices alone [13]. If this is the case, the question arises: How do customers decide from which e-retailer to purchase a product? This is the main question we address in this paper. Specifically, we seek to answer this question by applying statistical methods to a unique dataset on online prices and transactions collected from a Japanese price comparison site.

The rest of the paper is organized as follows. Section 2 provides a description of the dataset employed in this paper. Next, Section 3 first confirms the existence of substantial price dispersion on the price comparison website and then shows that customers choose the retailer from which they purchase a product based on the price rank rather than the price difference across retailers. Section 4 presents statistical regularities regarding at which price rank customers purchase, as well as at which price rank retailers post their prices when they enter the market. We show that both the probability of purchase by customers and the probability of price posting by retailers declines exponentially with price rank, and that the exponents associated with them are almost identical. This suggests that the reason why some shops set prices at a level substantially higher than the lowest price is that they know that some customers will choose them even at that high price. In Section 5, we then calculate the conditional probability that a retailer with a particular attribute (e.g., accepting credit card payment) is clicked on and compare this with the unconditional probability that a retailer is chosen in order to estimate the contribution of that attribute. Applying this idea, we estimate the brand value of shops. Finally, Section 6 concludes the paper.

## 2 Data

The data used in this paper are compiled from Kakaku.com, a major Japanese price comparison website [14], which lists product prices quoted by almost 2,000 consumer electronics retailers. (The number as of March 8, 2012, when we compiled our data, was 1,689.) Users of this website can find the prices quoted by retailers on the website as well as information on various retailer characteristics, such as whether they accept credit card payment, whether they also have physical retail premises, and the address of their distribution center. Consumers visiting the Kakaku.com website can use this information to choose a retailer from whom to purchase a product and can then click a button on the website that says “Go to retailer’s check-out page.” Our dataset consists of the records of all prices offered by each retailer (a total of around 802 million records) and the history of customer clicks on the “Go to retailer’s check-out page” button (around 210 million records) for all products offered from October 1, 2010 to January 31, 2012. In this paper, however, we focus only on the records for 6,385 major products that were sold for more than six months during this period and that received more than 1,000 clicks. The total number of clicks in connection with these products is about 110 million, constituting 50 percent of the total customer clicks during the observation period.

### 3 Price dispersion on Kakaku.com

Let us begin by examining price dispersion on Kakaku.com. The series denoted by  $\blacklozenge$  in Figure 1 shows the cumulative distribution of price quotes  $p$  relative to the lowest price  $\bar{p}$  for each product available at 0:00 on December 16, 2011. The tail of this distribution follows an exponential function of the form

$$P_{\geq}(\delta p) \propto e^{-\alpha \delta p}, \quad (1)$$

where  $\delta p$  is defined as  $\delta p = p/\bar{p}$  for each product and the estimate of the coefficient  $\alpha$  is 0.22. This figure shows that the fraction of retailers whose price quotes are more than 50 percent higher than the lowest price (i.e.,  $\delta p \geq 1.5$ ) is about 20 percent, clearly indicating the presence of wide price dispersion.

Next, we examine how customer clicks depend on the price gap between retailers. Specifically, we examine the relationship between the price gap between two retailers  $i$  and  $j$  of successive ranks (e.g., the first and the second, the second and the third, etc.), which is denoted as  $\Delta p_{i,j} = \frac{p_i - p_j}{p_j}$ , and the probability that retailer  $i$  will be clicked on, given that either  $i$  or  $j$  is clicked,  $P(i|\Delta p_{i,j})$ . The result is shown in Figure 2 and, not surprisingly, indicates that the probability  $P(i|\Delta p_{i,j})$  decreases the larger the price gap,  $\Delta p_{i,j}$ , between two consecutively ranked retailers. However, it is worth noting that the relationship between  $\Delta p_{i,j}$  and the probability that a retailer is clicked is discontinuous at  $\Delta p_{i,j} = 0$ . Specifically, when the price offered by retailer  $i$  is only 1 yen lower than the price offered by retailer  $j$ , retailer  $i$  is able to obtain 60 percent of the total clicks. However, even if retailer  $i$  continues to reduce the price and quotes a price that is 10 percent lower than that of retailer  $j$ , the fraction of clicks retailer  $i$  attracts increases only to about 70 percent. These results imply that customers choose a shop from which they purchase by focusing on the price rank gap between shops rather than on the simple price gap.

### 4 Customers' decision on where to purchase

In this section, we look at statistical regularities regarding at which price rank customers click on the ‘‘Go to retailer’s check-out page’’, as well as at which price rank retailers post their prices when they enter the market. Figure 3 shows the relationship between the price rank of a retailer and the probability that customers click on that retailer for a specific product, namely the Sony Blu-ray disc recorder with the model number ‘‘BDZ-AT700.’’ The figure indicates that although the retailer offering the lowest price attracts the largest number of clicks, this falls far short of an overwhelming majority, and that the click probability of the retailer offering the tenth lowest price shop is not zero. The click probability for the first-ranked retailer (offering the lowest price) is about 14 percent, that for the second-ranked retailer (offering the second lowest price) is about 11 percent, and that for the tenth-ranked retailer is about 3.3 percent. This probability distribution is well approximated by the exponential function

$$P_c(r) \propto e^{-a_c r} \quad \text{for } r \leq 25, \quad (2)$$

where  $P_c(r)$  is the probability of being clicked at rank  $r$ , and  $a_c$  is a coefficient, which is estimated to be 0.122.

We show that the relationship between price rank  $r$  and click probability  $P_c(r)$  of the large majority of products sold by more than 20 retailers follows an exponential function. To do so, we approximate for each product the click probability  $P_c(r)$  by a constant, a linear function, an exponential function, and a power law function, as follows:

$$\begin{aligned} P_c(r) &= c_0 \\ P_c(r) &= c_0 - c_1 r \\ P_c(r) &= c_0 e^{-c_1 r} \\ P_c(r) &= c_0 r^{-c_1} \end{aligned} \tag{3}$$

where the coefficients  $c_0$  and  $c_1$  are estimated using the maximum likelihood method. We compare these four specifications using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for each product and find that the exponential specification is chosen for 80.8 percent of all products, while the linear specification is chosen for 10.6 percent, and the power specification for 8.5 percent.

Next, we propose a hypothesis to explain the observed relationship between price rank and click probability. We focus on the difference in customer preferences regarding various retailer attributes. For instance, a customer who wants to pay by credit card will choose a shop that accepts credit card payment. We assume that a customer first chooses a set of retailers which satisfy a certain set of criteria determined by the customer, and then purchases the product from the retailer offering the lowest price among them. Importantly, customers are assumed to be heterogeneous in terms of their preferences over shop attributes. That is, some customers may prefer shops that accept credit cards, while others may not prefer such shops. Given these assumptions, the probability that a retailer with rank  $r$  in terms of price is clicked is given by

$$P_c(r) = \theta(1 - \theta)^{r-1}, \tag{4}$$

where  $\theta$  represents the probability that a particular retailer belongs to the set of favorite retailers for a customer. Equation 4 simply states that a retailer with rank  $r$  will be clicked only when none of the retailers offering a lower price are included in the set of favorite retailers. Comparing equations 2 and 4, we have  $\theta = 1 - e^{-a_c}$ . From this, we estimate that coefficient  $\theta$  is 0.115. That is, when 100 retailers sell this product, the number of favorite retailers is only  $100 \times 0.115 = 11.5$ . In other words, customers on average ignore 88.5 percent of retailers, including some or many that offer a lower price on the product the customer is interested in. Note that the coefficient  $a_c$  may differ across products. Figure 4 shows how the coefficient  $a_c$  for each product depends on the lowest price quoted for that product. The figure shows that there exists a convex relationship, with coefficient  $a_c$  highest for prices in the range of 10,000 yen (or about 100 US dollars), implying that customers do not pay much attention to shop attributes when they purchase products in this price range and price competition therefore is fiercer for such products.

Finally, we compare purchase prices (i.e, the price at which a customer clicked on the ‘‘Go to retailer’s check-out page’’ button) with sales prices (i.e., the price quoted by a retailer when it enters or re-enters the market) to make sure that price dispersion indeed stems from customers’ heterogeneous preferences over retailers. Note that, according to the rules set by Kakaku.com, retailers are not allowed to post prices when they have no inventory, so that retailers with no

inventory must exit from the market until they have the item in stock again. The sales price refers to the price quoted by retailers either at the time of newly entering the market or at the time of re-entering the market. The series denoted by ■ in Figure 3 shows how the probability that retailers post prices at rank  $r$  when they enter or re-enter the market depends on price rank  $r$ ,  $P_s(r)$ . The relationship is an exponential function of the form

$$P_s(r) \propto e^{-a_s r} \quad \text{for } r \leq 25, \quad (5)$$

which is similar to  $P_c(r)$  in equation 2. In fact, the exponents are  $a_s = 0.099$  for sales prices and  $a_c = 0.122$  for purchase prices and thus are quite close to each other. To check whether this result holds for other products, we compare  $P_s(r)$  and  $P_c(r)$  for all products sold by more than 20 retailers. The result is presented in Figure 5, which shows how the mean of  $P_s(r)$  depends on the value of  $P_c(r)$ . As shown in the figure, with a correlation coefficient of 0.65, these two probabilities are highly correlated, implying that retailers set a high price with a certain probability, because they recognize that customers click even at that high price with that probability.

## 5 Estimating retailers' brand value

In this section, we propose a method for estimating the brand value of a retailer by applying the line of reasoning regarding customers' choice of retailer discussed in the previous section. Let  $P_c(r|k)$  denote the probability that a retailer with a particular attribute  $k$  is clicked. We want to measure the value of this attribute. To do so, we employ the function  $B(k)$ , which is defined as follows:

$$B(k) \equiv \frac{1}{N} \sum_{r=1}^N \frac{P_c(r|k)}{P_c(r)}, \quad (6)$$

where  $P_c(r)$  is the unconditional probability given in equation 2, and  $N$  is the total number of retailers. Figure 6 presents the probability of being clicked for retailers that accept credit card payment, i.e.,  $P_c(r | k = \text{credit card payment})$ , showing that the probability declines exponentially with  $r$ , although the tail part deviates from a straight line. (We will come back to this issue later in this section.) Our estimate of  $B(k = \text{credit card payment})$  is 1.62, implying that the number of customers attracted by retailers accepting credit card payment is 1.62 times as large as the unconditional counterpart. We also find that  $B(k \neq \text{credit card payment})$  is 0.65, suggesting that retailers not accepting credit card payment attract 35 percent fewer customers than the average. We refer to  $B(k)$  as the brand value of a particular attribute  $k$ .

We apply this method to various retailer attributes and the results are presented in Tables 1 and 2. Table 1 shows the results for the availability of various payment methods. For example, in the case of the option to send cash via registered mail, the difference between  $B(k = \text{cash via registered mail})$  and  $B(k \neq \text{cash via registered mail})$  is very small, suggesting that it does not matter for consumers whether a retailer offers to accept cash via registered mail. However, for other payment methods, such as collect on delivery, bank transfer, payment by credit card, payment at convenience stores, and financing, it matters a lot for customers whether such a payment method is available or not. It should be emphasized that the number

of retailers accepting payment methods such as credit card payment, payment at convenience stores, or financing, is quite limited in this online market, as a result of which these retailers can attract more customers than other retailers.

Table 2 shows the result for the geographical location of retailers. One might think that it does not matter for customers where retailers are located, because customers do not actually visit the retail premises and shipping is free. However, the results presented in Table 2 show that the value of  $B(k)$  tends to be higher for retailers located in or near a major city like Tokyo, and lower for retailers located in prefectures far away from Tokyo. A possible reason is that customers may take into account the possibility that they have to visit the shop when serious problems arise.

As mentioned before, the tail part of  $P_c(r | k = \text{credit card payment})$  and  $P_c(r)$  in Figure 6 deviates upward from a straight line, which suggests that retailers with a large  $r$  may possess a number of attributes that are attractive to customers. In order to see whether this is true or not, we look at how the fraction of retailers accepting credit card payment is related to the price rank, which is shown by the series denoted by  $\blacklozenge$  in Figure 7. The figure suggests that retailers offering lower prices are less likely to accept credit cards. It can also be seen that the probability that a retailer accepts credit cards monotonically increases until the 15th price rank. Next, we repeat the exercise, but now change the definition of  $k$  to include a variety of payment methods, i.e.,  $k = \text{collect on delivery, bank transfer, payment by credit card, payment at convenience stores, and financing}$ . The result is shown by the series denoted by  $\blacksquare$  in Figure 7, which indicates again that retailers offering the lowest prices tend to not accept a wide variety of payment methods.

Finally, we estimate the brand value of each retailer by calculating the conditional probability  $P_c(r|i)$ , where  $i$  represents retailer  $i$ . The highest brand value among all the retailers,  $B(i) = 6.52$ , is recorded by a famous giant e-retailer known for offering a wide variety of products. We also find that some of the retailers with a high brand value are specialized in certain product categories such as wristwatches, air-conditioners, or in-car products. In contrast, shops with a small  $B(i)$  tend to be of small scale, and lack their own website and sell products only in online markets such as Yahoo, Amazon, and so on. In fact, the fraction of retailers without their own website is closely related with the value of  $B(i)$ ; that is, the fraction of retailers without their own website is 29 percent for shops with  $B(i) < 0.5$ , 7 percent for  $0.5 \leq B(i) < 1.0$ , and 1 percent for  $B(i) \geq 1.0$ .

## 6 Conclusion

In this paper, we established three empirical facts. First, we showed that prices quoted by retailers on a price comparison website, where search costs are negligible, show considerable dispersion. We also showed that customers click on the link to a retailer's website even if that retailer quotes a price that is substantially higher than the lowest price, although the probability that such a retailer's link is clicked is smaller than that for the retailer offering the lowest price. In other words, some shops quote higher prices in the knowledge that a fraction of customers will come even at those higher prices. Our second finding is that customers choose a retailer based on the price rank rather than the simple difference in quoted prices. For example, whether

a shop offers the first or the second lowest price matters, but the difference in yen between those two prices does not matter. Third, we showed that the probability that a retailer is clicked exponentially declines with the price rank and hypothesized that this stems from customers' heterogeneous preferences over a variety of retailer attributes. In fact, retailers accepting a wide variety of payment methods, such as credit card payment and collect on delivery, tend to attract more customers than retailers accepting a limited number of payment methods, and tend to sell products at higher prices. Based on these findings, we proposed a new method for quantifying the relative attractiveness of retailers, which we refer to as the brand value of retailers.

## Acknowledgments

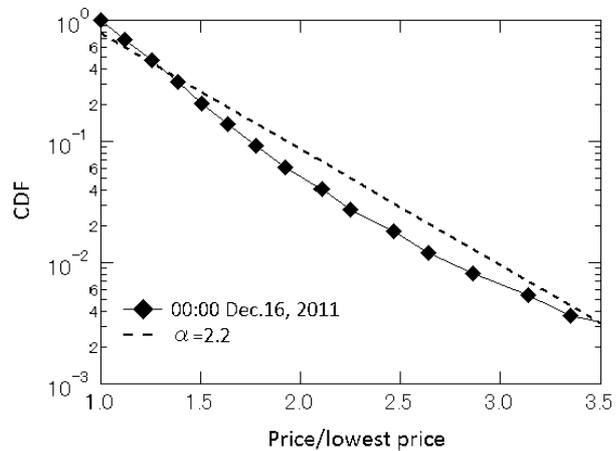
The data used in this study were obtained from the project on "Fostering Advanced ICT Professionals for the Age of Large-Scale Information Content" carried out at the University of Tsukuba with financial support from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT). This work was supported in part by a Grant-in-Aid for Young Scientists (B) (No. 24710156).

## References

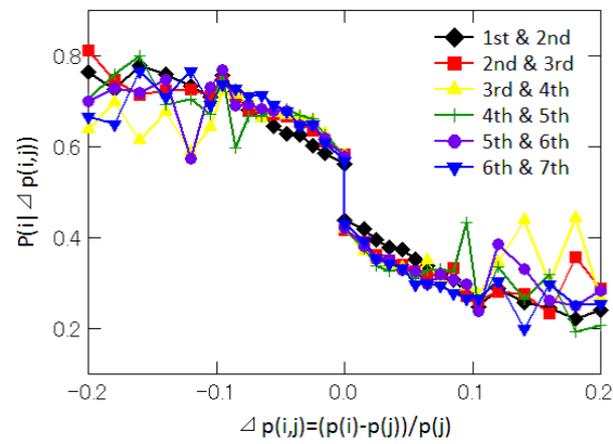
1. "Internet users per 100 inhabitants 2001-2011," International Telecommunications Union, Geneva. Retrieved 4 April 2012.
2. Bertrand J (1883) Review of "Theorie mathematique de la richesse sociale" and of "Recherches sur les principes mathematiques de la theorie des richesses." *Journal de Savants* 67: 499508.
3. *The Economist*, November 20th, 1990.
4. Bakos Y, Brynjolfsson E (2000) Bundling and Competition on the Internet. *Marketing Science* 19: 63-82.
5. Brown JR, Goolsbee A (2002) Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry. *Journal of Political Economy* 110: 481-507.
6. Lee HG, Lee SC, Kim HY, Lee RH (2003) Is the Internet Making Retail Transactions More Efficient? Comparison of Online and Offline CD Retail Markets. *Electronic Commerce Research and Applications* 2: 266-277.
7. Brynjolfsson E, Smith MD (2000) Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* 46: 563-585.
8. Smith M, Brynjolfsson E (2001) Customer Decision-Making at an Internet Shopbot: Brand Matters. *Journal of Industrial Economics* 49: 541-558.
9. Baye MR, Morgan J, Scholten P (2006) Information, Search, and Price Dispersion. *Handbook on Economics and Information Systems*.

10. Ellison G, Ellison S (2004) Search, Obfuscation, and Price Elasticities on the Internet. MIT Working Paper.
11. Baye MR, Morgan J, Scholten P (2004) Temporal Price Dispersion: Evidence from an Online Consumer Electronics Market. *Journal of Interactive Marketing* 18: 101-115.
12. Chevalier J, Goolsbee A (2003) Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics* 1: 203-222.
13. Ellison G, Ellison S (2005) Lessons about Markets from the Internet. *Journal of Economic Perspectives* 19: 139-158.
14. <http://kakaku.com/> (<http://corporate.kakaku.com/en/>)

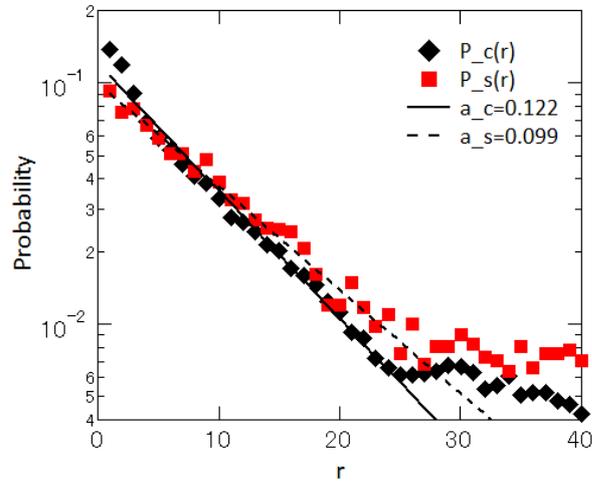
## Figure Legends



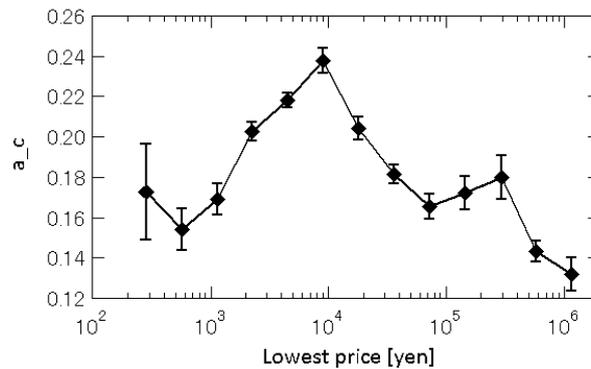
**Figure 1. Cumulative distributions of prices divided by the lowest price.** The series denoted by  $\blacklozenge$  shows the distribution of price quotes available at 0:00 on December 16, 2011, relative to the lowest price at that time. The dotted lines are reference lines representing an exponential function with an exponent of 2.2.



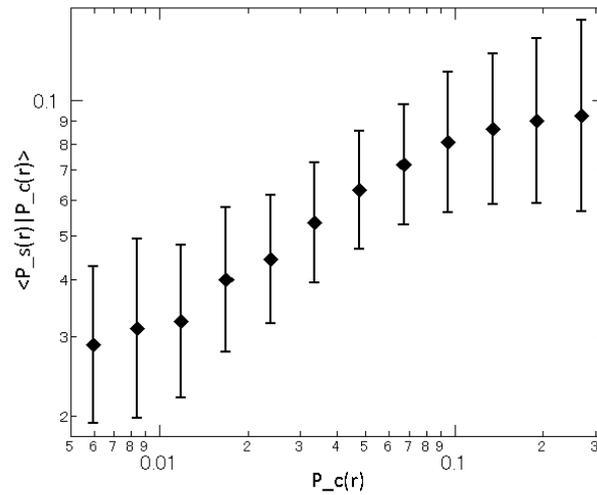
**Figure 2. Relationship between the price gap and the probability that a retailer is clicked.** The horizontal axis shows the price gap defined by  $\Delta p_{i,j} = \frac{p_i - p_j}{p_j}$ , where  $i$  and  $j$  are two adjacent numbers. The vertical axis shows the probability that retailer  $i$  is clicked on, given that either  $i$  or  $j$  is clicked, i.e.,  $P(i|\Delta p_{i,j})$ . The series denoted by  $\blacklozenge$ ,  $\blacksquare$ ,  $\blacktriangle$ ,  $+$ ,  $\bullet$ ,  $\blacktriangledown$  represent the results of the combination of the first and second rank, the second and third rank, the third and the fourth rank, the fourth and the fifth rank, the fifth and the sixth rank, and the sixth and the seventh rank, respectively.



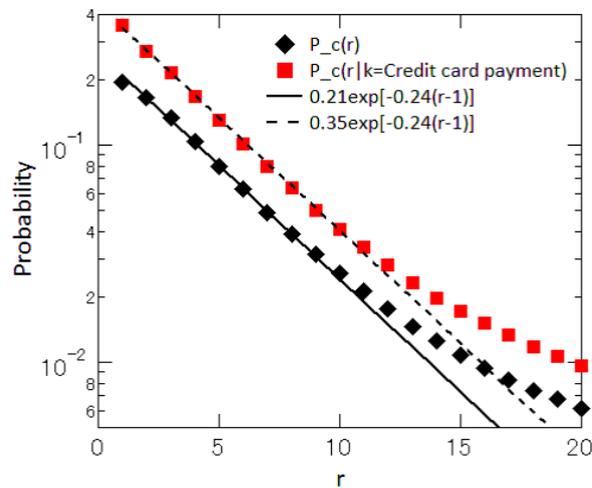
**Figure 3.** The probability that retailers post prices at rank  $r$  upon entry or re-entry to the market,  $P_s(r)$ , and the probability that customers click a retailer offering a price at rank  $r$ ,  $P_c(r)$ . This figure is for the Sony Blu-ray disc recorder with the model number “BDZ-AT700.” The solid and dotted lines represent exponential functions with exponents of  $a_c = 0.122$  and  $a_s = 0.099$ , respectively.



**Figure 4.** The estimated exponents  $a_c$  for different products. We split the entire sample of observed purchase prices into groups with different lowest prices at the time when clicks occurred, and then estimate  $a_c$  for each group.

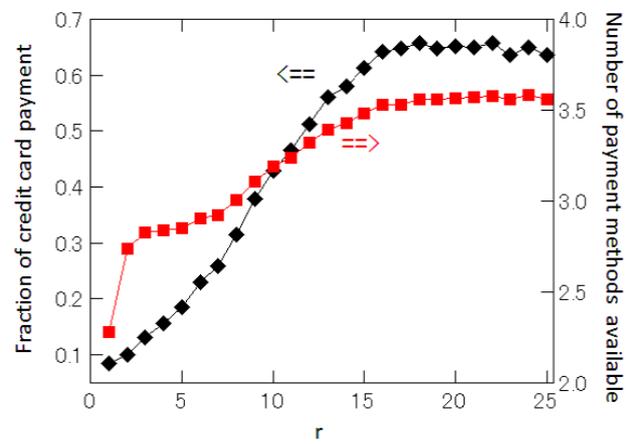


**Figure 5.** Relationship between the probability that customers click a retailer at rank  $r$ ,  $P_c(r)$ , and the probability that retailers post prices at rank  $r$ ,  $P_s(r)$ .



**Figure 6.** Probability of being clicked for shops accepting credit card payment.

The probability of being clicked for shops accepting credit card payment,  $P_c(r | k = \text{credit card payment})$ , is denoted by ■, while the unconditional probability,  $P_c(r)$ , is denoted by ◆. The solid and dotted lines are reference lines with an exponent of 0.24.



**Figure 7. Fraction of retailers that accept credit card payment, and the average number of payment methods available at each retailer.** The fraction of retailers that accept credit card payment is denoted by ◆, while the average number of payment methods available each retailer is denoted by ■.

## Tables

**Table 1. Estimates on brand value  $B(k)$**

Payment method $k$	Number of retailers	$B(k)$ for retailers where $k$ is available	$B(k)$ for retailers where $k$ is not available
Collect on delivery	1,589	1.027	0.799
Bank transfer	1,262	1.070	0.783
Credit card	895	1.617	0.649
Payment at convenience stores	324	1.335	0.937
Financing	217	1.277	0.909
Cash via registered mail	80	0.954	1.002

Note: The total number of retailers is 1,662.

**Table 2. Retailers' location and estimated brand value**

Ranking in terms of $B(k)$	Prefecture $k$	Number of shops in prefecture $k$	Travel time to Tokyo from prefecture $k$	Estimates on $B(k)$
1st	Gumma	21	1h16m	2.361
2nd	Kanagawa	96	0h32m	1.747
3rd	Ibaraki	11	1h56m	1.568
4th	Tokyo	556	0h00m	1.506
5th	Shizuoka	20	1h26m	1.302
29th	Ishikawa	10	3h30m	0.469
30th	Niigata	12	2h39m	0.409
31st	Nagasaki	5	4h00m	0.375
32nd	Yamaguchi	13	4h09m	0.313
33rd	Kagoshima	8	4h00m	0.288

Notes:  $B(k)$  is calculated for prefectures that have more than five retailers, which is the case for 33 prefectures. The table shows the top five and the bottom five prefectures in terms of the estimates for  $B(k)$ . The total number of retailers is 1,662.