NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Evgeny A. Antipov

THE DETERMINANTS OF ONLINE MERCHANT’S PRICE PREMIUM: EVIDENCE FROM RUSSIA

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: MANAGEMENT
WP BRP 19/MAN/2014

Evgeny A. Antipov[1]

THE DETERMINANTS OF ONLINE MERCHANT'S PRICE PREMIUM: EVIDENCE FROM RUSSIA[2]

Some Internet stores manage to charge prices that are significantly higher than market averages, therefore, obtaining some sort of price premium. This paper is dedicated to building a model that can be used to explain and predict a typical price premium that an Internet store charges for a specific product based on the information about the characteristics of the store and the features of the market for this product. Such models can provide support for pricing and assortment decisions: in particular, they allow detecting products that a store is likely to sell with the highest or the lowest markup based on price premia that are charged by stores with similar characteristics on similar markets.

JEL Classification: L81.

Keywords: hierarchical linear modeling, e-Commerce, price dispersion,

---

[1] National Research University Higher School of Economics (Saint-Petersburg, Russia). Department of Economics. Senior Lecturer; E-mail: eantipov@hse.ru

[2] This study was supported by The National Research University–Higher School of Economics' Academic Fund Program in 2013-2014 (research grant No. 12-01-0190)

1. Introduction

Empirical research have shown that there is persistent price dispersion even though price comparison can be easily done using various web services (Clemons, Hann, & Hitt, 2002). Recently it was shown that while retailer characteristics do impact online prices, this influence is significantly enhanced or diminished by the accompanying market characteristics (Venkatesan, Mehta, & Bapna, 2007), which is why hierarchical linear modeling (HLM) is a very promising methodology for explaining online price premium. To the best knowledge of the authors, the above-mentioned study by Venkatesan, Mehta and Bapna was the only one that explicitly incorporated interactions between retailer and market level factors. The study used data on the US online shopping. More studies that use newer data and/or data from other countries as well as modify sets of market and retailer characteristics would contribute to the empirical literature on explaining price differentiation online.

Besides being valuable for empirical generalizations in marketing science, hierarchical regression modeling can be used by marketers. For practical purposes it is useful for a merchant to learn what price premium to expect on a particular market. An expert system that uses information about the merchant i and about the market for product j as inputs and produces the expected price premium as an output looks extremely attractive for marketers: using such an expert system a merchant can offer more products that they can sell at a premium, while excluding some of the products with negative price premia from the assortment. It is worth mentioning that the data for such an expert system is publicly available in many countries thanks to the existence of price comparison websites. This fact significantly reduces the cost of practical implementation of such marketing information systems.

Our study explaining online price dispersion has several unique features:

- For robustness check we use 3 measures of price premium instead of a single one. These measures are described in the "Data" section of the paper
- We managed to account for such explanatory variables as whether the store has a quick order option at "Yandex Market" web service, whether it offers a warranty to the clients and whether it offers a credit to buy a product. What is more important, we are the first to account for the store website's citation index. By accounting for website quality we managed to test the hypothesis that follows from modern models of oligopolistic competition (Baye, Morgan, & Scholten, 2003; Varian, 1980). The hypotheses are detailed in Section 3 ("Hypotheses and model specification") of the paper.

2. Data

Our empirical analysis was based on 2584 price offers for 59 products from 3 product categories (smartphones, washing machines and refrigerators). The data was obtained from "Yandex Market" – the leading price comparison web service in Russia. It has many advantages compared to some of its Russian and international counterparts:

1. User-friendly interface that allows users to find the best offer easily. Some other websites use obfuscation strategies (Ellison & Ellison, 2009), i.e. make it difficult for users to search for the lowest price, so as to make merchants that offer high prices interested in being listed at the price comparison website.

2. The information about the prices and the availability is synced with the merchant's database. Merchants are penalized for providing wrong information by the quality assurance department of Yandex Corporation. In addition, customer reviews go through special fraud and spam filters. All these measures significantly improve data reliability.

3. High popularity of the web service among Internet shoppers (the number of unique visitors in 2013 was about 20 million people monthly[3]).

For robustness check we used 3 measures of price premium that merchant i charges for product j:

$$Price\_premium\_mean_{ij} = \frac{price_{ij} - price\_mean_j}{price\_mean_j} \tag{2.1}$$

$$Price\_premium\_median_{ij} = \frac{price_{ij} - price\_median_j}{price\_median_j} \tag{2.2}$$

$$Price\_premium\_min_{ij} = \frac{price_{ij} - price\_min_j}{price\_min_j} \tag{2.3}$$

The following explanatory variables were used:

- *quick_order* – a binary variable that equals 1 if the store has a "quick order" option that makes it possible for a user to fill in most of the fields in the order automatically with the information from her "Yandex Market" profile

- *warranty* – a binary variable that equals 1 if the store gives a warranty for the product

[3] http://stat.yandex.ru/

- *share_warranty* – the share of stores that offer warranty among all the stores that sell the product
- *price_mean* – mean price of the product across all the stores that sell it
- *rating* – store rating on "Yandex Market" (the number of stars from 1 to 5 based on customer reviews)
- *rating_mean* – mean rating of stores that sell the product (from 1 до 5)
- *rating_sd* – standard deviation of the rating of stores that sell the product
- *n_ratings* – the number of reviews for the store on "Yandex Market"
- *offline* – a binary variable that equals 1 if the store has both an Internet store and an offline store
- *CI* – the store's citation index that is used as a proxy measure for the quality of the website's SEO (search engine optimization)
- *credit* – a binary variable that equals 1 if the store offers a credit to buy a product
- *freedelivery* – a binary variable that equals 1 if the store offers free delivery
- *pickup* – a binary variable that equals 1 if the store offers Pick Up In-Store option
- *share_pickup* – the share of stores that offer Pick Up In-Store option among all the stores that sell the product

3. Hypotheses and HLM model specification

Equation 1:

$$\begin{aligned} Price\_premium_{ij} &= \beta_{0j} + \beta_{1j} rating_i + \beta_2 n\_ratings_i + \beta_{3j} pickup_i + \\ &+ \beta_4 offline_i + \beta_5 quick\_order_i + \beta_6 CI_i + \beta_7 noncash_i + \beta_8 credit_i + \beta_9 freedelivery_i + \\ &+ \beta_{10_j} warranty_i + u_{ij} \end{aligned} \tag{3.1}$$

Equation 1 allows testing the following hypotheses:

H1.1: The higher the i-th store's rating, the higher the price premium.

H1.2: The number of ratings positively influences the price premium.

H1.3: Pick Up In-Store option positively influences the price premium.

H1.4: Prices are higher in brick-and-mortar stores that also have an Internet store than in those stores that do their business only in the Internet.

H1.5: Stores that offer quick order option in association with Yandex have a higher price premium than stores that do not use this opportunity.

H1.6: The higher the citation index of the Internet store's website, the higher the price premium. This is our key hypothesis that was inferred from modern theoretical models of oligopolistic competition, according to which the informed customers (those, who use price comparison websites) pay the lowest price, while uninformed customers on average pay a higher price, because they do not use price comparison websites. An online store with high citation index is less dependent on buyers that come from price comparison websites, so having a well-optimized website should allow charging a higher price premium.

H1.7, H1.8, H1.9 and H1.10: The opportunities to pay with a plastic card, in credit, free delivery and warranty increase the price premium.

Equation 2: $\beta_{0j} = \alpha_0 + \alpha_1 n\_competitors_j + \varepsilon_{1j}$ (3.2)

The coefficient β_{0j} varies across markets depending on their competitiveness which is measured by the number of sellers that offer the j^{th} product.

H2.1: The number of competitors (n_competitors) that sell the j^{th} product negatively influences the price premium.

Equation 3:

$$\beta_{1j} = \gamma_0 + \gamma_1 n\_competitors_j + \gamma_2 price\_mean_j + \gamma_3 rating\_sd_j + \varepsilon_{2j} \quad (3.3)$$

H3.1: The positive influence of service quality on price diminishes as the number of competitors increases, because it may be increasingly difficult for shoppers to make optimal choices among a large number of merchants.

H3.2: The premium for store rating is higher for expensive products than for inexpensive ones, since buying an expensive product is risky.

H3.3: The marginal effect of service quality increases with the growth of service quality dispersion measured with its standard deviation. In the market where all sellers have approximately the same service quality high rating is a less significant advantage compared to that in a market where service quality is heterogeneous.

Equation 4: $\beta_{3j} = \delta_0 + \delta_1 share\_pickup_j + \varepsilon_{3j}$ (3.4)

H4.1: the positive effect of Pick Up In-Store option diminishes as the share of stores offering this option increases.

Equation 5: $\beta_{10j} = \lambda_0 + \lambda_1 share\_warranty_j + \varepsilon_{4j}$ (3.5)

H5.1: the positive effect of seller's warranty diminishes as the share of stores offering a warranty increases.

Although we could have added a few other equations to make the model even more flexible, we decided to leave it reasonably parsimonious. The single-equation model can be obtained by substituting equations 2-5 into equation 1:

$$\begin{aligned}
Price_premium_{ij} &= \left(\alpha_0 + \alpha_1 n_competitors_j + \varepsilon_{1j}\right) + \\
&\left(\gamma_0 + \gamma_1 n_competitors_j + \gamma_2 price_mean_j + \gamma_3 rating_sd_j + \varepsilon_{2j}\right) rating_i + \\
&+\beta_2 n_ratings_i + \left(\delta_0 + \delta_1 share_pickup_j + \varepsilon_{3j}\right) pickup_i + \beta_4 offline_i + \\
&+\beta_5 quick_order_i + \beta_6 CI_i + \beta_7 noncash_i + \beta_8 credit_i + \beta_9 freedelivery_i + \\
&+\left(\lambda_0 + \lambda_1 share_warranty_j + \varepsilon_{4j}\right) warranty_i + u_{ij}
\end{aligned} \tag{3.6}$$

Expanding the brackets results in the following equation:

$$\begin{aligned}
Price_premium_{ij} &= \alpha_0 + \alpha_1 n_competitors_j + \gamma_0 rating_i + \\
&+\beta_2 n_ratings_i + \delta_0 pickup_i + \beta_4 offline_i + \beta_5 quick_order_i + \beta_6 CI_i + \\
&+\beta_7 noncash_i + \beta_8 credit_i + \beta_9 freedelivery_i + \lambda_0 warranty_i + \\
&+\gamma_1 n_competitors_j \times rating_i + \gamma_2 price_mean_j \times rating_i + \gamma_3 rating_sd_j \times rating_i + \\
&+\delta_1 share_pickup_j \times pickup_i + \lambda_1 share_warranty_j \times warranty_i + \\
&+\varepsilon_{1j} + \varepsilon_{2j} rating_i + \varepsilon_{3j} pickup_i + \varepsilon_{4j} warranty_i + u_{ij}
\end{aligned} \tag{4.1}$$

4. Parameter estimates

The HLM model (equation 3.7) was estimated using restricted maximum likelihood (REML) implemented in SPSS IBM Statistics 20 (Table 1).

Table 1. Parameter estimates of HLM model's fixed effects (parameter estimates that are significant at 10% level are highlighted)

Parameter	Dependent variable								
	price_premium_mean			price_premium_median			price_premium_min		
	Estimate	Std. Error	Sig.	Estimate	Std. Error	Sig.	Estimate	Std. Error	Sig.
Intercept	1.419	2.515	0.573	3.417	2.832	0.228	7.829	3.509	0.026
n_competitors	-0.103	0.041	0.013	-0.113	0.048	0.020	0.118	0.060	0.051
n_ratings	-0.001	0.000	0.000	-0.001	0.000	0.000	-0.001	0.000	0.000
rating	1.079	0.645	0.095	0.891	0.703	0.206	-0.036	0.836	0.965
pickup	-1.799	1.708	0.293	-2.344	1.768	0.185	-6.241	2.256	0.007
offline	3.842	0.481	0.000	3.848	0.489	0.000	4.803	0.599	0.000
quickorder	-0.420	0.530	0.428	-0.435	0.539	0.420	-0.684	0.660	0.300

CI	0.002	0.000	0.000	0.002	0.000	0.000	0.002	0.000	0.000
noncash	0.434	0.953	0.649	0.522	0.971	0.591	1.031	1.188	0.386
credit	0.367	1.063	0.730	0.441	1.081	0.683	0.411	1.323	0.756
freedelivery	4.957	0.891	0.000	4.951	0.907	0.000	5.940	1.111	0.000
warranty	38.106	3.277	0.000	40.606	3.482	0.000	54.779	5.239	0.000
rating * price_mean	0.000	0.000	0.876	0.000	0.000	0.982	0.000	0.000	0.194
rating * rating_sd	-0.788	0.401	0.054	-0.580	0.499	0.249	0.865	0.580	0.139
pickup * share_pickup	-0.833	4.014	0.836	0.430	4.167	0.918	9.282	5.351	0.086
warranty * share_warranty	-59.649	5.597	0.000	-64.196	5.955	0.000	-87.323	8.917	0.000
n_competitors * rating	-0.003	0.011	0.813	-0.003	0.012	0.772	-0.013	0.013	0.304

We have found empirical support for our key hypothesis: the higher the citation index of an Internet store, the higher the price premium. From practical point of view this means that if an Internet store wants to avoid price competition, it should dedicate resources to SEO. The growing number of competitors decreases the price premium above the mean and median levels, but increases the price premium above the minimum level. A possible explanation is that when the number of competitors becomes larger, the probability of a discounter entering the market increases. The number of ratings has a negative impact on price premium, which is probably a result of reverse causality: sellers with lower prices are likely to have more reviewers. Seller’s reputation measured by its rating has a positive impact in price premium above the mean price, but not above the median or minimum price. This means that the effect of seller’s rating is somewhat ambiguous. Surprisingly, Pick Up In-Store option negatively influences price premium above the minimum and does not significantly impact other types of price premium. Offline stores are able to charge almost 4% more than pure-play online stores, which may indicate that people are ready to pay more if they know they can go to a physical store in case of a problem. Stores offering free delivery charge about 5% more than those that do not offer free shipping. The effect of warranty is generally positive, but diminishes when the share of stores offering warranty increases. Surprisingly, we have not found any increase in the effect of store ratings for expensive products compared to cheaper ones.

5. Conclusion

Based on an HLM-model explaining price premia of Internet stores on various markets, we have found empirical support for our key hypothesis: the higher the citation index of an Internet store, the higher the price premium. From practical point of view this means that online merchants that avoid price competition typically have better SEO.

Price premium modeling based on publicly available data has 2 main problems: the problem of endogeneity (low prices lead to higher ratings, not high ratings lead to low prices) and the lack of transactional data (we do not know whether anybody really buys a particular product from a particular seller at a particular price or not). The problem of endogeneity can be solved by using ratings of satisfaction with service quality, not with the price it charges. The elimination of price effects from user reviews of product or service quality is a problem deserving special attention in empirical market research. Researchers have shown that it is useful for web services to use multidimensional ratings to avoid biased ratings (Li & Hitt, 2010). Despite possible endogeneity problems and the lack of sales data the results are still appropriate for determining the size of price premium that different types of stores typically charge. Using market averages as a starting point for pricing policy is a common strategy for many stores.

References

1. Baye, M. R., Morgan, J., & Scholten, P. (2003). The value of information in an online consumer electronics market. *Journal of Public Policy and Marketing*, *22*(1), 17–25. doi:10.1509/jppm.22.1.17.17625
2. Clemons, E. K., Hann, I.-H., & Hitt, L. M. (2002). Price Dispersion and Differentiation in Online Travel: An Empirical Investigation. *Management Science*, *48*(4), 534–549. doi:10.1287/mnsc.48.4.534
3. Ellison, G., & Ellison, S. F. (2009). Search, Obfuscation, and Price Elasticities on the Internet. *Econometrica*, *77*(2), 427–452. doi:10.3982/ECTA5708
4. Li, X., & Hitt, L. (2010). Price Effects in Online Product Reviews: An Analytical Model and Empirical Analysis. *Management Information Systems Quarterly*, *34*(4), 809 – 831.
5. Varian, H. H. R. (1980). A model of sales. *The American Economic Review*, *70*(4), 651–659.
6. Venkatesan, R., Mehta, K., & Bapna, R. (2007). Do market characteristics impact the relationship between retailer characteristics and online prices? *Journal of Retailing*, *83*(3), 309–324. doi:10.1016/j.jretai.2006.04.002

Author:
Evgeny A. Antipov
National Research University Higher School of Economics (Saint-Petersburg, Russia). Department of Economics. Senior Lecturer;
E-mail: eantipov@hse.ru, Tel. +7 (906) 269-26-09

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.