PRICE DISPERSION ON THE INTERNET: EMPIRICAL COMPARISON OF SEVERAL COMMODITIES FROM THE CZECH REPUBLIC

Sedláček, J.

The first large-scale (909 products, 79,679 individual price listings) empirical study of price dispersion based on data from the Czech Republic's online shops is presented in the paper. First, simple descriptive indicators were calculated for each product. Second, several versions of linear regression models were constructed for each of the 13 product categories and evaluated against the hypotheses. The price dispersion was regressed against its mean market price and the number of shops. For the majority of categories R^2 is high or very high (the mean market price variable accounts for 86.6% to 96.7% in the price dispersion) or at least medium (from 56% to 69%). Also, price dispersion (measured by range and standard deviation) remains quite high for the majority of the analyzed product categories. The important role of pricebots for online shop owners and marketing managers and some other findings are also discussed.

Introduction

The nature of the Internet itself makes it relatively easy, fast and cheap for any consumer to compare prices in online shops. This feature is further enhanced by online price comparison services (shopping bots, price robots). The impact of Internet features on the different aspects of behavior of both consumers and online retailers has been investigated. Special attention has been given to the analysis of price dispersion (see the Short Review of Literature below for more information); however, most surveys were conducted in the U.S.A.

Therefore, the main goal of this paper is to carry out a study of price dispersion for selected product categories, based on empirical data from the Czech pricebot *Zboží*. Further questions are as follows:

- What pricebots are the most important for customers and retailers (or marketers)?

- Are a sufficient number of online shops in the Czech Republic registered on pricebots? (The evaluation will be based on a number of online shops listed on *Zboží* for particular goods and commodities.)

- Are prices of particular goods in some online shops "optimized" to achieve better positions in pricebot

listings? (This can be informally evaluated by looking for certain "patterns" in price distribution for particular goods.)

- How good of a source of data are shopbots for researches, retailers and marketing managers? Are there any particular differences or disadvantages?

H1: There is a positive relationship between the mean price of a standardized item sold on the Internet and price dispersion.

H2: There is a positive relationship between the number of available competitors of a standardized item sold on the Internet and price dispersion.

Short Review of Literature

On the Internet consumers can fairly easily navigate from one online retailer to another to compare prices, even without any assistance. This nature of the Internet was further greatly enhanced by online price comparison services (also called shopping bots, shopbots, price robots or pricebots). The first such service (BargainFinder) started soon after e-commerce was born in 1995 (for comparison, Amazon also opened in 1995.)

Some early articles therefore suggested that online shoppers would always search for bargain prices. It was assumed that there would be almost no price dispersion on the Internet. But it soon became evident that these assumptions were too simplified and one-sided.

Nowadays, pricebots offer much more functionality than the so-called 1st generation pricebots (like BargainFinder). However, the only two indisputable conclusions are:

- The Internet has greatly reduced the costs to consumers of identifying the sellers of various goods and the prices that they charge.

- Nowadays many consumers often use shopbots to search for information when making purchase decisions on the Internet.

It is not so clear and consistent what the other consequences of online shopping are; a lot of articles and studies have been published. We will mention only a few of them. Lindsey-Mullikin and Grewal (2006) empirically test the notion that the degree of dispersion increases with the mean price. Two studies were conducted: the first for TV and VCRs, the second for PDAs, laptops and MP3s (also more non-price measures were evaluated: enabled by improved data set provided by a shopbot). The results of both studies were consistent: the Internet has not commoditized products. In addition, the number of retailers, the type of retail stores, and the quality rating of the stores influence the degree of market price variation.

Nelson, Cohen, Rasmussen (2007): in the first part of the article there is a good summary of many previous studies on the topic. While earlier surveys of price dispersion have produced mixed results, they have confirmed the notion that price dispersion on the Internet is far from ceasing to exist. Nelson, Cohen and Rasmussen continue with their own research. While previous studies had one or two commodities, this survey made a comparison of 542 products across 13 categories. The main results are: the level of price dispersion is positively related to the price of the product and the number of sellers, and lower for goods that would typically be purchased several times a year. Other factors are shipping costs or seller heterogeneity.

Another group of researchers therefore has concentrated on explanation of why price dispersion continues to exist in online shopping. The main factors identified in several studies include seller characteristics, number of sellers and time of entry. Ellison and Ellison (e.g. 2004, 2009) identified another, previously unmentioned, explanation about search, obfuscation and price elasticity.

The major reason why the large study by Chu, Chintagunta & Cebollada (2008) is so interesting is comparison of price sensitivity both for online and offline channels. Among the conclusions are: nearly 90% of households shop both at online and offline stores. Across 12 very different product categories the price sensitivity is lower when shopping online than offline. The authors also suggest several explanations for this.

A study by Brynjolfsson, Dick & Smith (2010) comes to several interesting conclusions. First, they estimate the costs of search and benefits (both using a major shopbot) for books. A rather surprising suggestion is that consumers face significant search costs, even in a "nearly perfect" market. Price elasticity is relatively high compared to offline markets. Furthermore, contrary to the common assumption, search intensity is not correlated with greater price sensitivity. Instead, consumers who search multiple screens put relatively more weight on non-price factors like brand.

Allen and Wu (2010) took another approach to investigate the role of shopbots in Internet shopping. In an extremely large-scale study (2.2 million vendor price listings from eight shopbots) they examine how accurately shopbots represent a market and also analyze the strategies shopbots adopt to achieve market representativeness.

Their main conclusions are: (1) shopbots do not represent markets equally well; (2) size drives a shopbot's market representativeness positively whereas affiliation drives a shopbot's market representativeness negatively; (3) shopbots follow different vendor representative strategies to pursue market representativeness.

Even from a short review of the literature above, it is clear that a lot of articles and studies were published on the topic of pricing on the Internet in general and about price dispersion in particular. However, most of these studies are from the U.S. market.

As far as we know, there are just two similar empirical studies from the Czech Republic and both with a limited number of goods. Our own study (Sedláček, 2012) was conducted as a pilot study with just one commodity: LED TVs. Finally, we had 216 HTML files, 65 TV models (only TV models with no data from online shops or data from just one shop were eliminated) and 3,737 individual prices. A similar study for another commodity (mobile phones) has been performed by a student as part of his diploma thesis (Pešek, 2012).

In the pilot study we have verified in practice all important parts of the research, e.g. which shopbot(s) we should choose, how many shops have a price listing for the given product, how difficult or easy is it to download and extract all data, etc.

Methods

Therefore, the research presented in this paper is a logical continuation of the above mentioned pilot study. However, this time we have conducted full scale research. Originally,

Table 1: Basic quantitative characteristics for each commodity Number of products: how many different products (models) are in each product category.

Product category	Number of		Online shops		Average price		Price range	
	prod.	prices	min	max	min	max	min	max
Hard disks (HDD)	80	16594	26	670	1131	33320	392	12009
Lenses for DSLR	74	7314	41	130	4028	57784	1992	32566
GPS navigations	59	5073	25	130	1769	22576	505	6993
Laser Printers	50	6231	78	171	2285	163582	466	53918
IPads (Apple)	32	4222	37	175	8925	23792	1859	7116
DLP projectors	36	4176	50	142	5617	51183	1402	36256
LCDs	36	3606	33	139	2068	15072	389	7165
Flash disks	43	8354	51	328	169	19129	314	17335
Digital SLR (DSLR)	29	3245	66	138	12101	71052	3830	31474
Graphics cards	56	5485	28	159	659	23851	237	18744
Perfumes	135	4498	6	77	297	2789	70	2162
Sleeping bags	33	1238	17	80	677	7283	86	2197
Tires	246	9643	3	90	1045	11185	214	7179
Total	909	79679	—	—	—	—	—	—

Number of prices: how many individual price listings are in each product category.
Online shops: the smallest (the largest) number of online shops having the particular product.
Average price: for the cheapest (the most expensive) product in each product category.
Price range: the smallest (the largest) price range (max price – min price) for the particular product (not necessarily for the cheapest or the most expensive product).
Notes: Only selected descriptive characteristics are presented in table to it will remain uncluttered. However, for price dispersion both range and standard deviation were always calculated and both versions of indicators were used in regression models (see later). For comparison with the surveys mentioned in the Literature Review, average price (and not median) was used in regression models.
Source: author, data from Zboží (2013)

data for 10 commodities were collected. Later, we added three more product categories to further diversify the selection of products. Finally, we have 13 commodities, 909 products, (different product models) and 79,679 individual price listing. Thus, the sample size is larger or at least comparable to a vast majority of the studies mentioned above in the literature review (e.g. the study by Nelson, Cohen and Rasmussen had 13 categories, but only 542 products). Of course, the study by Allen and Wu is the exception, but the main goal of that study is quite different.

Selection of Pricebot

Online price comparison services do not have as long of a tradition in the Czech Republic as elsewhere in the world. Nevertheless, there are quite a few shopbots in our country. But we can easily say currently there are only two that really matter: Zboží and Heuréka.

Zboží (2013) gains a great portion of its popularity from the fact that it is part of Seznam (still the most visited web portal and search engine in the Czech Republic). Seznam had more than 5.6 million unique visitors per month in December 2012, according to NetMonitor (2012). Zboží itself had about 2 million unique visitors per month in December 2012, again according to NetMonitor. The Zboží server (almost) in its current form was established in 2007; some additional functions were added in 2009.

Heuréka (2013), as the major competitor to Zboží, was also established in 2007 (in beta version). It had more than 2.6 million unique visitors per month in December 2012, again according to NetMonitor (2012). This number suggests that Heuréka is even bigger than Zboží; however, the number of visitors is not the best metric for such websites. We think that the number of finished purchases would be much better, but no such figures are available.

The market share of all remaining shopbots of Czech origin is small. However, in the future we should pay attention to Google Shopping. In the Czech Republic, Google started the transition of Google Shopping to the monetized service in February 2013. We can expect that this service will be the major competitor to Zboží and Heuréka (similar to the

Figure 1: Comparison of predictor variables (product category: hard disks)
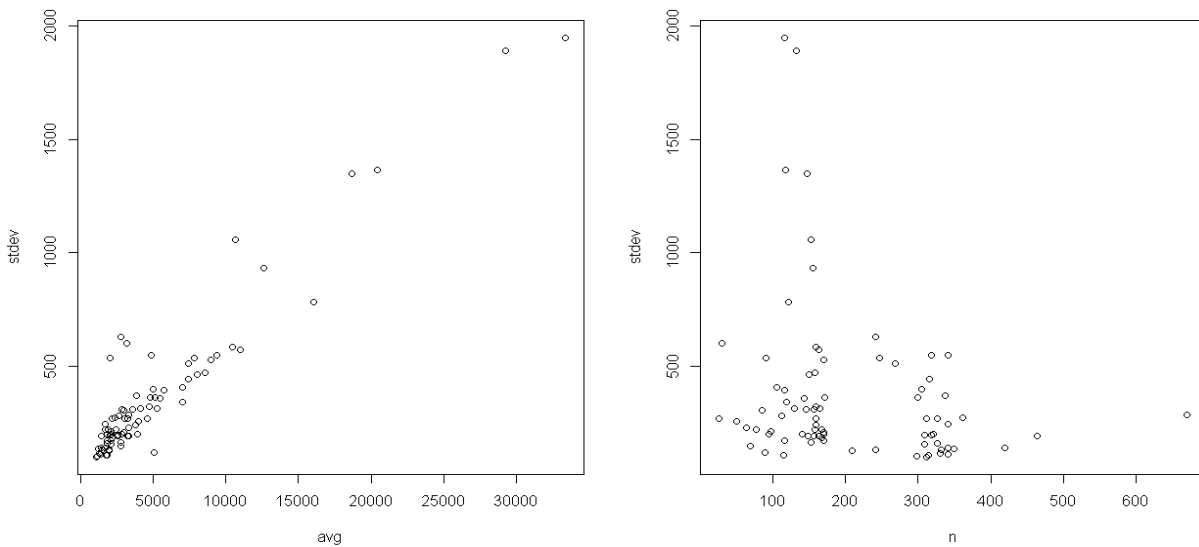
already existing Google – Seznam situation for general search services).

We have not succeeded to obtain data directly from Zboží or Heuréka. Therefore, we have decided to use data just from one of the two main shopbots (from Zboží), mainly because it is quite difficult and time consuming to download such data.

Data Gathering

This section gives some rather technical detail to ensure reproducibility of the process of data collection and of the whole research. Download of HTML pages: unfortunately URLs at Zboží (or at Heuréka) are sometimes quite inconsistent, even for similar products. Therefore, for each product we had to retrieve at least one HTML page manually (using WWW browser as a normal user). The URL and the number of subsequent pages for the given model were recorded. Quite often, the search had to be modified to filter out unwanted items (like accessories) and alternate URL was recorded. Then, for each category all pages were downloaded again (semi-automatically using GNU Wget), usually on Saturday or Sunday or overnight, to minimize the possible changes in the database before the download was completed (Zboží also puts a quite strong limit on the number of files we can download in one hour from one IP address). In total, it took several days to download all data.

Selection of relevant data: In the second phase, individual prices were extracted from HTML pages (using GNU grep) and consolidated into one file (for each commodity). We already knew from the pilot study that consumers' rating for online shops (1 to 5 stars) is quite common, but not available for all shops. Therefore, we did not take this variable

into account, but it may be useful for some future research. Statistical analysis: First, simple descriptive indicators like average, max, min, median or standard deviation were calculated. Second, several versions of linear regression models were constructed and evaluated against hypotheses. All these calculations were performed using R (R Core Team, 2012).

Results

Some results (simple descriptive indicators for each category) have already been presented in Table 1. Here we will concentrate on regression analyses. For each product category we consider two predictor variables: the average price of each product (avg) and the number of online shops having a particular product (n). As already mentioned, we use avg (and not median) mainly for better comparison with surveys mentioned in the Literature Review. The dependent variable is price dispersion: measured by price range (only minimal and maximal prices are taken into account) or by standard deviation (stdev). For the first product category results for all versions of multiple and single regression models are shortly presented and compared in Table 2.

For the first model (range ~ avg + n in Wilkinson-Rogers notation) and first product category, both predictive variables passed statistics tests (p-values or t-test). However, the test values are quite different and, if we look at a simplified model (range ~ avg), we can see that R^2 is not very different. This may indicate that the number of online shops is not a very good predictor variable. This can be further confirmed by some other analyses or simply by drawing range (or stdev) against avg and range (or stdev) against n. (See Figure 1 for version with stdev; all graphs were created, but they are not shown due to space limitations).

Now we can try to express price dispersion by the standard deviation method instead of range. Thus, model No. 3 is stdev ~ avg + n. At first glance, the results are similar to model No. 1, but this time, variable n did not pass tests (p-value is too high; t-test is too low). Therefore, we will drop n from the model and model No. 4 is just stdev ~ avg. And this is "the best" model for the first product category. We can clearly see that the coefficient of determination (R^2) is high (about 90%) and the model is significant for any reasonable significance level.

Table 2: Comparison of all regression models for first product category (hard disks)

a. Hard disks: Multiple Linear Regression Model (range ~ avg + n)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	83.1057	274.7288	0.30	0.763
avg	0.3534	0.0198	17.83	< 2e-16
n	3.4799	0.9914	3.51	0.00075

Residual standard error: 965 on 77 degrees of freedom
Multiple R-squared: 0.806,   Adjusted R-squared: 0.801
F-statistic: 160 on 2 and 77 DF, p-value: <2e-16

b. Hard disks: Simple Linear Regression Model (range ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	900.0806	156.1795	5.76	1.6e-07
avg	0.3349	0.0204	16.38	< 2e-16

Residual standard error: 1030 on 78 degrees of freedom
Multiple R-squared: 0.775,   Adjusted R-squared: 0.772
F-statistic: 268 on 1 and 78 DF, p-value: <2e-16

c. Hard disks: Multiple Linear Regression Model (stdev ~ avg + n)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	48.99691	31.80342	1.54	0.13
avg	0.05932	0.00229	25.85	<2e-16
n	0.08283	0.11477	0.72	0.47

Residual standard error: 112 on 77 degrees of freedom
Multiple R-squared: 0.902,   Adjusted R-squared: 0.9
F-statistic: 355 on 2 and 77 DF, p-value: <2e-16

d. Hard disks: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	68.44263	16.84334	4.06	0.00011
avg	0.05888	0.00221	26.70	< 2e-16

Residual standard error: 111 on 78 degrees of freedom
Multiple R-squared: 0.901,   Adjusted R-squared: 0.9
F-statistic: 713 on 1 and 78 DF, p-value: <2e-16

Source: author, data from Zboží (2013)

For all other goods, again all four versions of the models were constructed, but only the results for the "best model" are presented. For 10 of 13 categories, the best model is *stdev ~ avg*. For tires and perfumes *range ~ avg* is slightly better. In six cases, R^2 is high or even very high (from 86.6% to 96.7%), for four categories R^2 is let's say "medium" but still good (from 56% to 69%) and only for other two categories R^2 is relatively low (35% and 40%). The IPads category is different. In this case "the best model" is *stdev ~ avg + n* with R^2 about 78.5% (for comparison, *stdev ~ avg* has R^2 about 61%).

Table 3: "The best models" for other categories

Lenses for DSLR: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-37.1517	66.9937	-0.55	0.58
avg	0.0793	0.0028	28.32	<2e-16

Residual standard error: 329 on 72 degrees of freedom
Multiple R-squared: 0.918,   Adjusted R-squared: 0.916
F-statistic: 802 on 1 and 72 DF, p-value: <2e-16

GPS navigations: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.57e+02	8.09e+01	1.94	0.057
avg	4.44e-02	8.02e-03	5.54	8e-07

Residual standard error: 266 on 57 degrees of freedom
Multiple R-squared: 0.35,   Adjusted R-squared: 0.339
F-statistic: 30.7 on 1 and 57 DF, p-value: 8.01e-07

Laser printers: Simple Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-46.58854	69.62639	-0.67	0.51
avg	0.04971	0.00133	37.50	<2e-16

Residual standard error: 354 on 48 degrees of freedom
Multiple R-squared: 0.967,   Adjusted R-squared: 0.966
F-statistic: 1.41e+03 on 1 and 48 DF, p-value: <2e-16

Apple IPads: Multiple Linear Regression Model (stdev ~ avg + n)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-251.1181	108.6453	-2.31	0.028
avg	0.0466	0.0046	10.13	4.9e-11
n	2.1940	0.4493	4.88	3.5e-05

Residual standard error: 98.9 on 29 degrees of freedom
Multiple R-squared: 0.785,   Adjusted R-squared: 0.77
F-statistic: 52.8 on 2 and 29 DF, p-value: 2.16e-10

DLP projectors: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.33e+02	1.54e+02	0.86	0.39
avg	5.13e-02	6.96e-03	7.37	1.5e-08
Residual standard error: 482 on 34 degrees of freedom Multiple R-squared: 0.615, Adjusted R-squared: 0.603 F-statistic: 54.3 on 1 and 34 DF, p-value: 1.54e-08				

LCDs: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-59.75868	25.46408	-2.35	0.025
avg	0.06400	0.00431	14.85	<2e-16
Residual standard error: 80.6 on 34 degrees of freedom Multiple R-squared: 0.866, Adjusted R-squared: 0.862 F-statistic: 221 on 1 and 34 DF, p-value: <2e-16				

Flash disks: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.56e+02	3.59e+01	4.34	9.2e-05
avg	9.38e-02	9.84e-03	9.53	5.8e-12
Residual standard error: 206 on 41 degrees of freedom Multiple R-squared: 0.689, Adjusted R-squared: 0.682 F-statistic: 90.9 on 1 and 41 DF, p-value: 5.84e-12				

Digital SLR: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.80e+02	8.14e+01	-2.22	0.035
avg	8.00e-02	2.88e-03	27.76	<2e-16
Residual standard error: 222 on 27 degrees of freedom Multiple R-squared: 0.966, Adjusted R-squared: 0.965 F-statistic: 771 on 1 and 27 DF, p-value: <2e-16				

Graphics Cards: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.39824	66.51150	0.17	0.86
avg	0.05855	0.00969	6.04	1.5e-07
Residual standard error: 328 on 54 degrees of freedom Multiple R-squared: 0.403, Adjusted R-squared: 0.392 F-statistic: 36.5 on 1 and 54 DF, p-value: 1.45e-07				

Perfumes: Simple Linear Regression Model (range ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	30.6902	56.8610	0.54	0.59
avg	0.6301	0.0462	13.65	<2e-16
Residual standard error: 263 on 133 degrees of freedom Multiple R-squared: 0.584, Adjusted R-squared: 0.58 F-statistic: 186 on 1 and 133 DF, p-value: <2e-16				

Sleeping Bags: Simple Linear Regression Model (stdev ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-20.61814	17.87692	-1.15	0.26
avg	0.08836	0.00871	10.15	2.3e-11
Residual standard error: 59.9 on 31 degrees of freedom Multiple R-squared: 0.769, Adjusted R-squared: 0.761 F-statistic: 103 on 1 and 31 DF, p-value: 2.25e-11				

Tires: Simple Linear Regression Model (range ~ avg)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-652.5257	142.6989	-4.57	7.7e-06
avg	0.6392	0.0365	17.52	< 2e-16
Residual standard error: 813 on 244 degrees of freedom Multiple R-squared: 0.557, Adjusted R-squared: 0.555 F-statistic: 307 on 1 and 244 DF, p-value: <2e-16				

Source: author, data from Zboží (2013)

The results indicate that a significant portion of price dispersion for a given commodity of goods sold on the Czech Internet can by explained according to hypothesis one. The higher the average price is, the greater the price dispersion is (and for most commodities, R^2 is at least medium, quite often high or even very high).

On the other hand, we cannot confirm hypothesis two: there seems to be some relationship between the number of online shops and price dispersion. But it often seems to be very different for individual products and the global values are either very weak or even insignificant. If we look at our findings and also at the literature, probably the best explanation is collision of two competing theories. According to one theory, with more shops there is stronger competition (and price dispersion should decrease). On the other hand, each shop has a slightly different price and general marketing strategy (and price dispersion should increase). In practice, it is very unclear, if and which factor prevails.

How can the different results for IPads be explained? First, this product category is quite homogenous, but not as homogenous as many readers may expect. (There are three product lines: IPad Mini, IPad 2 and IPad Retina. Within each product line, there are several models with different memory, etc.) Second, IPads are not only high-tech, but also high-touch products, they are in great demand and online shops do not want to sell them at a very "discounted" price as is common with many other products. Third, there are big differences in the number of online shops offering individual models: some are available only from about 37 to 39 shops, others are in 130 to 175 shops (and almost nothing in between). Fourth, IPad and Apple are premium brands. This

is certainly reflected in the Apple's brand and price policy and (indirect) influence of Apple on dealers.

Unfortunately, we do not have clear explanations why GPS navigation and graphics cards are the two categories with quite low R^2. This can be investigated in some future research. On the other hand, there are quite convincing arguments why, for the majority of categories, standard deviation gives better results in regression models than range. Range (by definition) takes only maximum and minimum into account and thus can be much more affected by outliers than standard deviation.

Managerial Implications

First of all, the presented results clearly show that price dispersion remains quite high for analyzed goods from Czech online shops (despite easy comparison of prices online, strong competition of shops and price sensitivity of Czech customers). And for the majority of analyzed commodities, price dispersion is strongly related to the average price.

Besides "the hard facts" from statistical analyses, we can find some other, less formal conclusions based on non-aggregated data, descriptive statistics and on the whole experience gained during the preparation of this paper. (The data are too big to be presented in paper.)

Shopbots do not have such a long tradition in our country compared to the U.S.A, but their increasing popularity among customers can be demonstrated in several ways, not only by the number of visitors (e.g. data from Netmonitor). Currently, the two most important shopbots in the Czech Republic are Zboží and Heuréka. In the future, Google Shopping (relatively new in our country) can also gain significant market share. While no direct data are given in the paper, shopbots are now an important source of visitors (and buyers) for many online shops. Every marketer or retailer should pay great attention to this fact.

Many online retailers already utilize shopbots in several ways. Again, in addition to other numbers, we can note supporting findings in distribution of individual prices, e.g. quite often the lowest prices differ only by 1 CZK or so. Then there is a gap. It can be random sometimes. But this pattern is quite frequent in the collected data; therefore we think we can attribute it to careful selection of price for given goods. In other word, prices of many goods in some online shops are already "optimized" to achieve better positions in the pricebot's listing (and thus received more visitors or even buyers). Finally, we can evaluate shopbots as a source of data for research (and marketers, too). Can we find a sufficient number of online shops in the Czech Republic on Zboží (or Heuréka) and for the commodity we are interested in? Of course, for any given product, we will never get data from all Czech online shops. Some shops are still not registered (on Zboží or Heuréka or both); others have decided not to

show data for the given model in pricebots, for some reason. However, if we look at Table 1, the maximum numbers of shops are typically between 77 and 175 shops, in one case 328 and once even 670. Of course, for some other commodities (not analyzed in the paper), the number is lower. However, we can say that for a lot of categories, the sample size we can obtain this way is sufficient. And (if necessary) data from a shopbot can be supplemented by data collected directly from selected online shops (either in a similar fashion like the method described in this paper or using some commercial, paid service).

In general, Zboží (or Heuréka) contains a tremendous number of useful marketing data and several types of analyses can be conducted (even if some data like consumer ratings are not available for all shops). We can analyze, e.g. the development of price (min, max, average) over some time for selected goods, compare price strategy of a large online shop to the whole market in different segments, etc.

However, shopbots in our country have also some disadvantages from this aspect. As already mentioned, download of a larger set of data is cumbersome and time consuming (due to inconsistency of URLs and strong limits on downloads).

The last disadvantage is not specific to shopbots in the Czech Republic. We have no information about the real number of items sold for any particular price. And we can only guess that for the lowest prices there will certainly be some purchases and possibly no purchases for the highest ones.

References

Allen, G.; Wu, J. (2010). How well do shopbots represent online markets? A study of shopbots vendor coverage strategy. *European Journal of Information Systems*, 19(3), 257–272.

Brynjolfsson, E.; Dick, A. A; Smith, M. D. (2010). A nearly perfect market? *Quantitative Marketing and Economics*, 8(1), 1–33.

Ellison, G.; Ellison, S. F. (2004). Search, obfuscation, and price elasticities on the internet. *National Bureau of Economic Research*, Working Paper 10570, [available at www.nber.org/papers/w10570].

Ellison, G.; Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2), 427.

Heuréka (2013) (accessed February 10, 2013), [available at http://www.heureka.cz/]

Chu, J.; Chintagunta, P.; Cebollada, J. (2008). A comparison of within-household price sensitivity across online and offline channels. *Marketing Science*, 27(2), 283–299, 308.

Lindsey-Mullikin, J.; Grewal, D. (2006). Imperfect information: The persistence of price dispersion on the web. *Academy of Marketing Science Journal*, 34(2), 236–243.

Nelson, R. A.; Cohen, R., Rasmussen, F. R. (2007). An analysis of pricing strategy and price dispersion on the internet. *Eastern Economic Journal*, 33(1), 95–110.

Netmonitor: veřejné výstupy (2012). (Měsíční zpráva – Prosinec 2012) (accessed February 10, 2013), [available at http://www.netmonitor.cz/verejne-vystupy].

Pešek, Tomáš (2012). Cenová politika na Internetu se zaměřením na cenové agregátory. Diploma Thesis, Prague: University of Economics, Prague, 2012

R Core Team (2012). R: *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Sedláček, J. (2012). Pricing on the Internet: Empirical Study from the Czech Republic. In: Černá, I. (ed.). *Czech Republic and Slovakia in International Trade and Business 2012*. Prague : Oeconomica Publishing House, 2012, pp. 129–143.

Zboží.cz (2013) (accessed from February to March, 2013), [http://www.zboží.cz/]

Author

Ing. Jiří Sedláček, Ph.D.

Senior lecturer

Department of International Trade

Faculty of International Relations

University of Economics, Prague

nám. W. Churchilla 4, 130 67 Praha 3, Czech Republic

sedlacek@vse.cz