

Analysis & visualization of price dispersion

Product - Mobile speakers

Authors Name: Rahul Chaturvedi
e-mail: v20rahch@du.se

Authors Name :Firas Fourati
e-mail: h20firfo@du.se

Abstract—The clearing house model implies that retailers will try to appeal to the two types of customers in the market (shoppers and loyal). However, the success of this model will rely on an important component which is Intertemporal price discrimination. In this paper we tried to investigate the reason behind price dispersion for the Mobile speaker's product. After using Time's series clustering, we concluded that price dispersion is caused by intertemporal price discrimination strategies followed by some retailers.

Keywords-component; Price dispersion; Intertemporal price discrimination ; clustering ; price comparison

I. INTRODUCTION

The successful availability of price comparison websites has led the market stakeholders to question the fundamental issue on how market functions in terms of price differences. Retailers and regulators are interested to understand based on the clearing house models either retailers are reducing and increasing their prices intertemporally to stay competitive in the market or are they following one pricing strategy [4]. In this research paper, the analysis of price dispersion for one category of electronics products "Mobile Speakers" is performed and the attempt to visualise the data as well. The findings will have implications either for retailers to charge higher prices and earn extra profits or for the favour of customer who can learn which shop is expensive and avoid it.[4]

II. LITERATURE REVIEW

Many researchers tried to investigate the issue of price dispersion and reasons behind it. Stigler 1961 [1] mentioned that there is an obvious price dispersion within similar categories of one product in developed countries. Varian ,1980 [2] was the first to emphasize the importance of the clearinghouse model which is one of the fundamental models used to understand the pricing behaviours of retailers. This model implies a retailer will try to attract together price sensitive customers "shoppers" and non-price sensitive customers "loyals". This model has put the conditions for the

price dispersion to remain as follow: First, existence of loyal customers. Second, retailers appealing to the two types of customers and finally, retailers randomizing their prices [3]. This approach will lead to the so-called intertemporal price discrimination situation which will not be easily figure out by customers. However, the nonexistence of this approach will result in consumers buying only from cheaper retailers, and expensive retailers will either disappear or they will follow low price strategy only [4]. C. lindgren et al, 2020 tried to investigate if intertemporal price discrimination the cause of price dispersion in markets with low search costs. His results show that clustered retailers remain in the same price clusters over time even after checking for retailer's heterogeneity. According to the author, price dispersion is not explained by intertemporal price discrimination strategies. Our research will try to answer the same questions investigated by previous research for one specific category of electronics product.

III. METHDOLOGY AND DATA COLLECTION

A. Data collection & exploration

The data was retrieved from a price comparison website called Price spy Sweden for the years 2012 until early 2017. Importing the dataset revealed 12027209 observations and 14 variables. Description of the dataset' variables was provided in addition with the data. It contains prices of retailers for each product-id for each specific date. Other important variables are included like CPI adjusted price to account for inflation and log of CPI adjusted price to account for variabilities. We were able to get a grouped view of the dataset using the different methods in R-studio. The grouping showed that there is 197 product-id and 253 store-id. The summary statistics showed that the number of products in the market has increased significantly starting from 2014 and price differences are very important between low and high price products, ranging from 29 SEK until 38583 SEK.

B. Methodology for analysis

R programming and packages were used for the analysis and visualizations of the dataset provided. Due to the

existence of 197 product-id and 253 stores and to be able to analyse the price's dispersion, we plotted the prices for five product-id separately that we think they make a summary of the whole dataset. the products-id chosen contains either the minimum, maximum, 1st quartile, 3rd quartile and median's prices of the mean prices of the products.

To perform the clustering, Kabacoff, R. 2015 [6] emphasized to follow certain steps which we considered to understand the principles of clustering and advance further in our analysis.

The purpose to arrange the prices along with timeseries in different clusters with different pricing level was done with help of clustering. Once clustering is done, it became very easy to present the movement of price series across the clusters. The clustering was performed on product level, not category level to maintain the sanity of the cluster for a particular product.

Although scaling of the prices was proposed before clustering, which is supposed to help in better view of price dispersion, but it was observed that results were not much different using scaling. Hence the prices were kept intact during clustering to ensure that the readers get the good graph which are easy to interpret with the linear scale of CPI adjusted prices.

For clustering the timeseries within the products, the alignment of the prices with timeseries was required beforehand. We used dynamic time warping (dtw) method to calculate the distance matrix to ensure that stores selling the product in different time duration to be clustered further, however, other methods of clustering could also be used to calculate the distance.

Next step was to select a clustering algorithm. Hierarchical clustering was chosen to divide the stores timeseries to prepare clusters and analyse the movement of stores along the time duration from one price range to another. It was decided that 3 clusters would be enough to represent the result in the best way, hence it was selected manually instead of deciding the same via algorithm.

IV. RESULTS AND FINDINGS

A. Descriptive statistics

Fig. 1 represents summary statistics of prices for all products-id in the form of boxplots. It includes the mean in red dots, median, 1st quartile and 3rd quartile, and minimum values along with the outliers in the graph. Starting from 2013 the mean is almost twice greater the median, which means that the distribution is positively skewed [5].

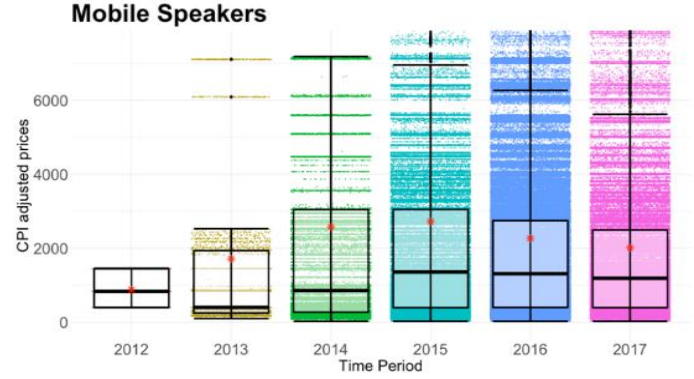


Figure 1. Summary statistics at category level

B. Pricing behaviour of retailers

Fig. 2 represents the product-id which has the minimum's average price in the mobile speaker's market with 121,5 SEK sold by 13 stores. It is visible that few stores following low price strategies over time, and others do not have a stable pricing strategy. For example, we can see clearly that store with colour in pink has moved from low to medium pricing strategies more than five times.

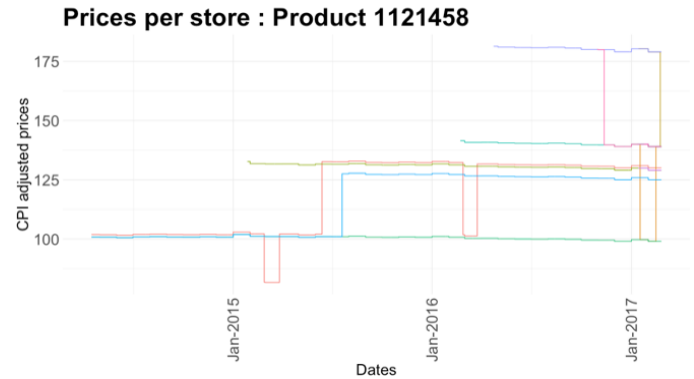


Figure 2: Product with the min mean price of mobile speakers.

Fig. 3 represents the product-id that has the max's average's price in the mobile speaker's market with 31069 SEK, sold by 5 stores. We can see there are stores following stable low-price strategy. over time, others are randomizing their pricing strategies, which means they are discriminating price sensitive consumers when they decided to increase the price.



Figure 3: Product id with the max mean price of mobile speakers.

Fig. 4 represents the product-id that has the 1st quartile average's price of the whole dataset sold by 20 stores. We can see a few stores following a stable low-price strategy. However, the other stores have started with medium pricing strategy then moved to higher pricing strategy. This is another case that clearly shows a price movement between times which is discriminating a proportion of the consumers. Retailers starting with low prices then moving to higher ones in this case may be explained as well by a strategy to start with low prices as they are selling this product for the first time.



Figure 4: Product-id representing the 1st quartile of mean price of mobile speakers.

Fig. 5 represents the product-id that has the 3rd quartile average's price. this product model is sold by 20 Stores who are following a stable and unique strategy over time (either low, medium, or high). In such a case, the changes in pricing is not important enough to say that these product-id retailers are following an intertemporal price discrimination.



Figure 5: Product-id representing the 3rd quartile of mean price of mobile speakers

Fig. 6 represents the product-id that has the median of average's price sold by 21 stores. We can see that there are a few stores following low price strategies. The Other retailers clearly following medium price strategy only. There is a third category of customers who are sometimes lowering prices and other times raising prices which means that they are trying to make more profits by attracting non sensitive customers only.

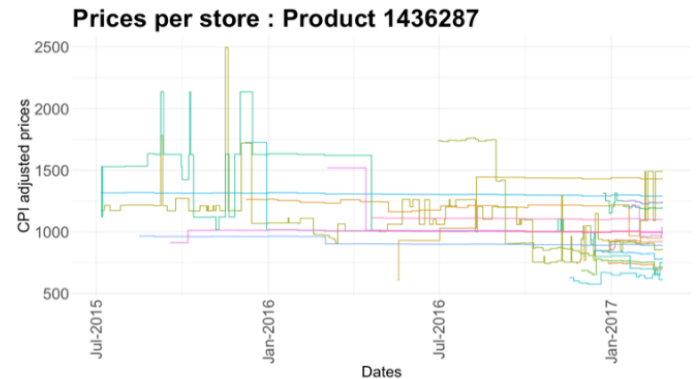


Figure 6: Product-id representing the median of average price of mobile speakers.

From the summary statistics on the pricing behaviours of retailers inside each product-id, we can see that in 4 out of 5 product-id retailers are following an intertemporal price discrimination. However, to have a more inclusive conclusions about the status of the market we need to make a time series cluster analysis.

C. Cluster analysis

Cluster Dendrogram : Product ID 1121458

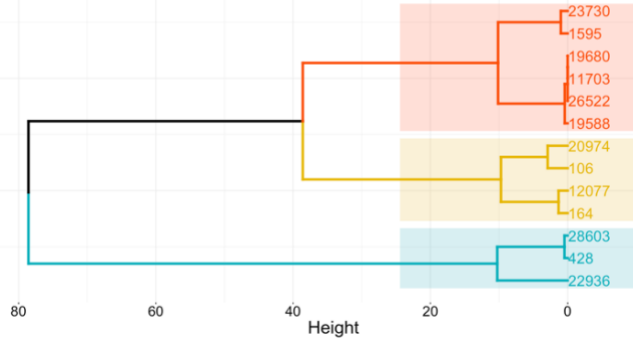


Figure 7: Complete-linkage clustering of store_ids

Fig. 7 shows how the store-ids are clustered to three groups for the product-id 1121458. The goal here is to arrange the prices along with timeseries in different clusters. The method of clustering is complete linkage here.

In this section, to satisfy the heterogeneity of the analysis and because of the huge number of products-id, we are going to present four examples of clustering before we conclude with a summary about the whole dataset.

In fig. 8, we can see clearly that cluster 1, composed of 1 retailer is making an obvious intertemporal price discrimination over time. This retailer was charging less than 500 SEK in the first three month of the year 2016. Then he started to charge almost 1500 for a period of one year, then he reduced the price to around 1000 SEK.

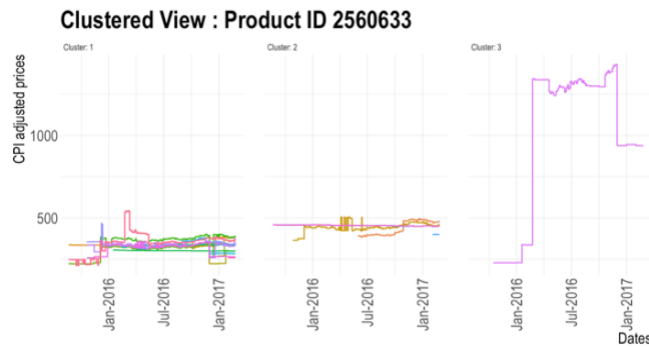


Figure 8: Pricing strategy of 13 stores selling 2560633 Product.

Fig. 9 shows the clusters for the product id that has the minimum average's price. The only remark here is that cluster 1 has few retailers who are randomizing their price from one time to another which shows that they are trying to attract sensitive price shoppers sometimes and eliminating them in other times.

Clustered View : Product ID 1121458

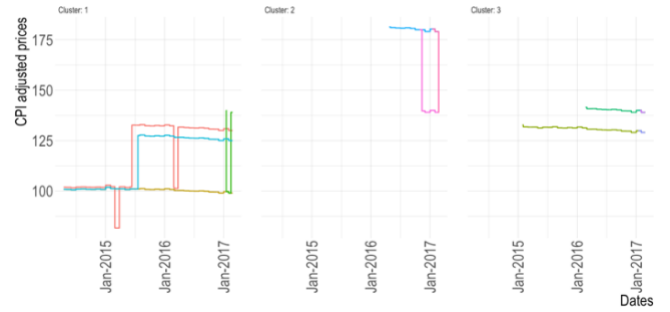


Figure 9: Pricing strategy of 13 stores selling 1121458 Product.

Fig. 10 shows a price dispersion for a very short duration. In such a case we cannot conclude either this dispersion is caused by an intertemporal price discrimination or by a normal pricing strategy.

Clustered View : Product ID 1557082

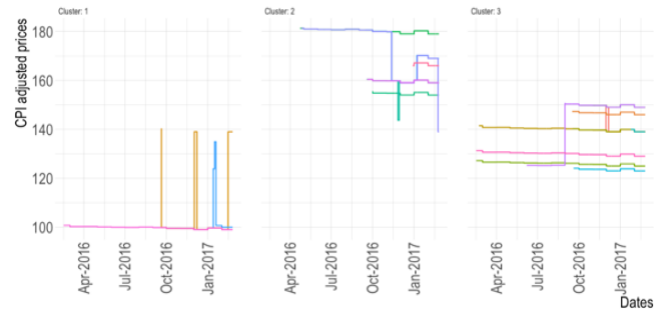


Figure 10: Pricing strategy of 20 stores selling 1557082 Product.

Fig. 11 shows the clusters for the product that has the maximum's average price. we can see that cluster 1 & 2 are trying to attract sensitive price shoppers sometimes and eliminating them other times to make extra profits, through randomizing the price.

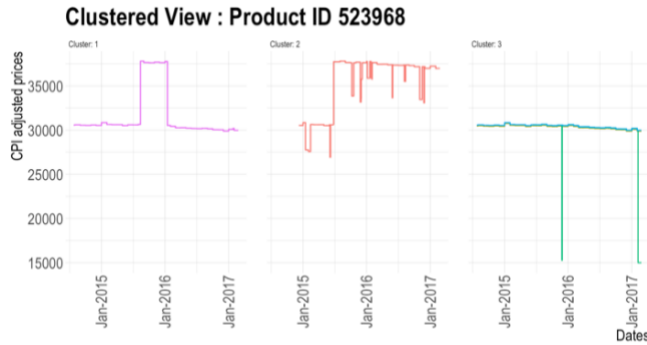


Figure 11: Pricing strategy of 5 stores selling 523968 Product

We further tried to visually examine the price dispersion for 11 product-id as a part of data exploration and conclude that 9 of them contains stores in clusters which were randomizing their price multiple times, which is also a sign of intertemporal price discrimination. So overall, we realise that total 12 clusters out of 15 are following the strategy of intertemporal price discrimination.

V. CONCLUSION

In this paper, the issue of price dispersion was investigated using the techniques of clustering and data visualizations. Using summary statistics for the category level revealed a clear price dispersion on the market. Further, we used the hierarchical clustering to arrange the prices along with time series in different clusters and different pricing level. This led to confirmation that approximately 80% of the retailers 'clusters we examined are following intertemporal price discrimination which is causing price dispersion. With the power of data visualizations techniques applied in this research we can say that the findings are considered an opportunity for retailers to increase prices (randomly) and earn extra profits.

ACKNOWLEDGMENT

We would like to thank Lars Rönnegård and Charlie Lindgren for their untiring support in explaining the requirements of the project as well as detailed description for preparing the report. Also, the pieces of code provided by Charlie for clustering part were of great help as well.

REFERENCES

- [1] Stigler, G. J. "The Economics of Information." *The Journal of Political Economy* 69 (3), 1961: pp.213–225. doi:10.1086/258464.
- [2] Varian, H. R. "A Model of Sales." *The American Economic Review* 70 (4), 1980: pp.651–659.
- [3] Lach, S. "Existence and Persistence of Price Dispersion: An Empirical Analysis." *Review of Economics and Statistics* 84 (3), 2002: pp.433–444. doi:10.1162/003465302320259457.
- [4] c. lindgren et al. "Is intertemporal price discrimination the cause of price dispersion in markets with low search costs?", 2020. *Applied Economics letters* <https://doi.org/10.1080/13504851.2020.1789055>
- [5] Lumen learning, "Skewness and the Mean, Median, and Mode", <https://courses.lumenlearning.com/introstats1/chapter/skewness-and-the-mean-median-and-mode/>
- [6] Kabacoff, R. R in action: Data analysis and graphics with R (2.th ed.). Manning, 2015, pp.369-382