Essays in platform economics

Paolo Marocco

Department of Economics (DIEC)

University of Genoa

Supervisor

Anna Bottasso, University of Genoa

In partial fulfillment of the requirements for the degree
of
PhD program in Economics

December 4, 2020

Abstract

The Multi Side Marketplace (MSM) is an invention of internet era. The MSM arises from a confluence of different services and portals, which started the Web revolution. When compared with the village market and the traditional brick-and-mortar store, the platform offers the capability to integrate different market sectors in a unique paradigm. For example, Amazon, Facebook, Uber, Netflix and other companies create value through different business models involving several requirements, oriented towards more than merely satisfying consumer needs. Indeed, they scrape and collect a large amount of information for learning the market opportunities beyond selling a good, for releasing collateral frameworks, far from main tasks (such as a tool to design web components or Content Management System).
Generally speaking, online platforms are matchmakers. David S. Evans and Richard Schmalensee[1] describe such platforms as facilitating good matches between two sides: one that sells and another that buys. This is simple when the match involves two agents, but can became complex when are more. To this end, it is useful to show an example reported by Rysman about externalities generated by the main process: "In the case of a video game system, the intermediary is the console producer—Sony in the scenario above—while the two sets of agents are consumers and video game developers. Neither consumers nor game developers will be interested in the PlayStation if the other party is not"[2]. The multi-side nature of the platform must control the mutual interactions between parties, not only to increase mutual benefits, but also in the case in which they have neutral or adverse purposes. The capability to match a very large number of agents (sellers, buyers, social media users, advertisers, software developers, etc.), under the constraint to control the compliance, is often a search problem. Contrary to village markets, platforms use matching algorithms facing toward this goal. The search algorithm formalizes an issue that only a single user was capable of previously in retailing: reducing time and increasing the best match quality. However, the search algorithms digest information and need a large quantity of information in order to match well. The more information they have characteristics of regarding the agents and goods, the better the search ranking and the lower the search costs for all. Platforms such as Google, Amazon, Facebook continue to ingest more and more information at an ever increasing rate, but this operation has a

[1] Evans, David S., and Richard Schmalensee. Matchmakers: The new economics of multisided platforms. Harvard Business Review Press, 2016.

[2] Rysman, Marc. "The economics of two-sided markets." Journal of economic perspectives 23.3 (2009): 125-43.

cost. Platforms have by-passed this issue, promising benefit for the consumer, in exchange for uploading free contents, so-called User Generated Contents (UGC) without making explicit what was implicit. For example, the UGC are often compared with the Public good (non-excludable and non-rivalrous[3]) without a compliant model of public good (eg.: conceptualized by a game or a behavioral experiment). Although it does not completely reach this goal, the novelty of our work proceeds through a formal approach to decode the informative signals of UGC, to demonstrate their usefulness in product search. This regards both achieving more product variety in less time and exploiting a decision making mechanism to help the consumer best match their taste. In details, each chapter is focused on the quality-price relationship. The first one investigates the concept of product quality, trying to capture its endogenous qualities and convert them into an index. The second exploits this index toward searching and ranking processes, formulating a decision mechanism able to suggest the optimal choice. The last one analyzes a competition between sellers with respect to the price proposed through a dropping price tactic.
We present the work *Product quality of platform markets*, in which we propose a product quality index based on UGC, whose insights have been captured by means of Machine Learning and Natural Language Processing framework. This framework has been customized to set a non-trivial improvement in information analysis from economics point of view. The aim of this index is to drive consumers' choices for finding trustworthy signals of actual quality, enhancing a timeless response to their searches. We have shown how our quality index can improve product price differentiation, applying our model to a sample of electronic camera products. The paper was written with Michela Chessa (Université Côte d'Azur). The second work, *Optimizing Product Quality in online Search*, is the result of a collaboration between myself, Anna Bottasso (Università di Genova) and Michela Chessa (Université Côte d'Azur). In this article we have described a model of consumer sequential search for products, exploiting a decision making mechanism based on a Stopping Rule. We show how the consumer can optimize their search strategy, demonstrating an improvement in terms of consumer utility at different levels of price, with respect to a rating-based search mechanism. The original aspect of the work is the adoption of a Statistical Learning algorithm, to define a new endogenous concept of quality, set as the core of the searching process. The final chapter, a collaboration between myself, Anna Bottasso (Università di Genova) and Simone Robbiano (Università di Genova), is *Price Matching and Platform Pricing*. The study refers to the price strategy where different retailers commit themselves to match any lower price offered by competitors on the same item or the product category. While such guarantees are widespread policies among retailers, in literature there is no consensus on the view that low-price guarantees are used to effectively discourage price cutting, but rather they

[3][Users cannot be barred from accessing and/or using them for failing to pay for them. Meaning that their use by one person does not affect others' ability to use them, and users cannot be prohibited from their use by failure to pay for them.

encourage anti-competitive behaviors. In the paper we empirically investigate the effects of price matching guarantees on U.S. consumer electronics online marketplace, by means of a unique dataset developed through sophisticated and computerized scraping procedures.

In sum, in this thesis we present three papers which investigate informative content generated by consumers, aiming to improve the usefulness for matching high quality products at lower prices. Following a general perspective, we explore platform product listing, searchable through a decision making mechanism. In a more specialized perspective, we take into account a dropping price modality service, differentiating the consumer benefit in the case of high or low quality product matching.

Contents

List of Figures

List of Tables

0 Chapter 1

Product quality on platform markets

Abstract

Many studies have questioned the meaning of "product quality", hanging between a characteristic interpretation of a product for improving consumer satisfaction, and scientific approach to measure its benefits. Starting from the historical quality setting as mirror image of the price, we investigate the adoption of new signals, developed over the years to adjust the original relationship. Recently, bootstrapping by emperor of e-commerce platforms, the rating system has emerged as a reference contribute for product quality informativeness. We study this tendency, to show its failure in the presence of low price market and new brands. For this purpose, we collect User Generated Contents from a well-known online retailing platform. We capture and distill meaningful features in order to adjust the rating assigned by reviewers, and propose a novel quality formula able to increase the accuracy of the information provided to the consumer. We suggest that our formula better captures product quality, and, when adopted by a platform for sorting the products, it increases the products variety and, consequently the satisfaction of the consumer. Our proposal suggests a way to facilitate the consumer search (as we will show in the second chapter). Moreover, it can be used as a measure of market efficiency in the case of voluntary opacity of the platform in exposing product quality signals.

1 Introduction

Price and quality are the main pivotal indices that play a relevant role in product managing. Historically, most of the problems in estimating the quality arise from the very real difficulty of even reaching a consensual meaning of the word "quality", for which there are a multitude of definitions. An historical product quality definition is that approved by the International Standards Organization (ISO). According to this definition, quality is *the totality of features and characteristics of the product or service that bear on its ability to satisfy stated or implied needs revolving around customer*[4]. This definition focuses the attention on quality as a key attribute of some product features, as they are discovered by the consumer, such as design, conformance, and other parameters of fitness for use. "Consumer Reports", a magazine which

[4]Mentioned in The ISO 8402-1986 standard protocol. ISO is an international standard-setting body composed of representatives from various national standards organizations, founded in 1947.

represents the point of reference for testing products[5], has standardized and shared the information about these features since 1936. This magazine has become a point of reference for consumers and scholars whose interest is in the correspondence between price and quality, producing a rich statistical source of multi-dimensional indices. This information has also played a crucial role in discovering violations of federal laws and norms, such as for the scandal related to the Volkswagen in 2011[6]. With the rise of Web platforms, a much larger information set about product features has started to be easily at disposal of every consumer. The phenomenon of the influence of reviewers' opinions on consumer choice is nowadays present on a much larger scale, thanks to the so-called *User Generated Contents (UGC)*. Our work tackles the issue of proposing a new way to measure quality, i.e., a new quality index, which is based on UGC. Our index inherits the current and common perspective to adopt the rating as a basic signal for quality, but complements it by taking into account other key variables that can be extracted from the large amount of information available online. We investigate information that is made available by the online platforms and define some key features. First, the already mentioned rating. Secondly, the number of likes (Thumb Up/Down), also called votes, which evaluate the trustability of the reviews and the reputation of the writer. The votes represent the response of a large public to accept/deny the rating decided by a small population of reviewers. Other signals extracted from the text content evaluate when the votes are representative of the review, in particular when they infer the reviewer reputation (eg.: the length and density of the review). Taking quality as the simple mirror image of the price, scholars in the seventies developed studies to analyze other variables. In the eighties, the simple alignment price/quality questioned, extracting the tangible components whose quality was the manifestation of labor, design, durability, marketing investments and so forth. The advertising expenditure played an increasingly important role in the nineties, when shopping has landed on line, and grew up the capability to track different signals, mixing ADV cost, price and effective purchasing. In the new millennium, the upsurge of large-scale operators such as the platforms of retailing, has upgraded the opportunities to find out product information and to dive deeply in the details (big/small brands, variety of market categories, and so forth), and consumer perceptions (thrustability and consensus). The digital economy opened the doors to the so called *Big Data* analysis, without worrying anymore about how to manage all the information. It became one of the main tasks of researchers in digital economics to investigate this "mare magnum" to reach useful fragments of hid-

[5] Consumer Reports (CR) is an American nonprofit consumer organization dedicated to unbiased product testing, investigative journalism, consumer-oriented research, and consumer advocacy.

[6] At the starting of controversy involving Volkswagen not regular emission mode, there has been a test of Consumer Report in 2011 (focused on the car "Jetta SportWagen TDI"). Volkswagen admitted to using the defeat device, and has been ordered to recall approximately 482,000 cars. United States federal penalties have been severe, including fines ranging up to US $18 billion.

den information, able to reveal market and product insights. Our work renews the canonical adoption of product features, as we will discuss in the related literature, but in a completely new format, as the confluence of economics and other branches of science. We adopt processes involving the capture of heterogeneous information (like quantitative values and linguistic text) that typically concerned computer science research, that only recently have been the instruments of economics analysis. The text analysis is the principal task of Natural Language Processing (NLP) tools, which transform linguistic information to a vector space, offering the possibilities to merge and integrate values coming from different kinds of data sources. Using these tools we have the possibility to discover, create and collect different product attributes in order to create a function able to explain a trustable quality index. For example, we have calculated the relevance of the reviewer's reputation, through the estimation of the link between votes and text reviews'. Also, we have discovered that a review's length is another important factor for the consumer's attention, but up to a certain threshold, above which the reading is interrupted. This paper aims to answer the following research question: how is it possible to adjust a canonical quality index, such as the rating, submitting it to a revision of trustability? Our solution suggests a method based on public endorsement. We capture the best features are able to settle the study just explained, obtained through a feature selection mechanism endogenous to the quality index computation. We provide results about the relationship between the price and our quality index, and we offer guidelines which are useful to evaluate the efficiency of the market and to set new policies for public intervention, especially in the case of voluntary opacity of platforms in exposing quality information. The remaining of this paper is organized as follows. The following Section 2 reviews the literature on product quality, also drafting around the collection and distillation of the UGC ingredients, retrievable on the platforms. In Section 3 we present what a UGC is in detail, and why and how it can be useful for a quality evaluation, in particular on the inefficiency of the rating system. Section 4 presents the data collection, the features extraction and descriptive statistics. In Section 5 we present how our detailed reviews data are handled through statistical learning algorithms, in order to select the main important features and to define the new product quality formula. Section 6 concerns our quality index, bringing out the experimental evidence and discussing the empirical results, comparing a selection which is sorted by rating with a new one which is sorted based on our quality index. Section 7 concludes. The details about features list and statistical learning sections have been moved to Appendices.

2 Related literature

The difficulty meaning of "quality", gave rise to doubts about the uniqueness of the definition. For example, Holbrook and Corfman[24], in the eighty, explored the literature about the concept of quality in various disciplines. The disparity among the different definitions collected by the authors, suggests

the obstacles to a conceptualization and reflect a failure to distinguish quality from other types of values as beauty, convenience, and fun. Concerning the consumer expectations, the question reflects linguistic performance in the actual usage of the term to describe consumption experiences. Taking a similar direction, Hjorth-Andersen[26], in the same years, analyzes the difference between mono and multi-dimensional index of quality, in order to capture the inefficiency of the consumer market clear representation of the price as an indicator of quality. The author compares works aligned to price as quality, with other ones in opposition. The conclusions show the relevance of the relationship price-quality, sustained by a richer quantity of multi-dimensional index data, extended to more categories. Fundamental for the research is the use of the dataset of product features of "Consumer Reports Magazine"[7]. Curry and Riesz[12] try to trace a correspondence of price and quality over time, here too, using empirical data reported in "Consumer Reports magazine" from January 1961 through December 1980. Unlike the research of Hjorth-Andersen, focused on the number of features and the complexity of information to define product properties, Curry and Riesz adopt as guidelines three theoretical hypotheses about product pricing policy: PLC (product life cycle) theory, dynamic pricing policy, and economic information theory. Through this model, the authors capture certain structural demand characteristics that directly influence consumer behavior and indirectly producer strategy. The results show a compliance the PLC theory, in which prices converge to an optimal level of price and quality over the years, as well as decrease in price variation, apparently due to a close level of prices by competitors.
In the nineties, the correspondence between price and quality was criticized aggressively by different authors, among them Philip Crosby [11], who has explored in which way different processes of manufacturing and distribution can change the price, maintaining the same quality level. The new perspectives have overpassed the previous approach, taking into account a relationship between production cost and prediction of demand, focusing attention brand competition, return of investment and capability to evaluate product impact. All these issues conveyed product quality information on marketing, in particular the advertising power to affect the consumer decisions. The perspective, formalized by Milgrom and Roberts [36], is that sellers of high quality goods will spend more on advertising because they will benefit from an increment of potentially happy buyers. The decontextualized pair price-quality therefore entails a correction, due to the quality perceived by the user under the influence of advertising. This assumption is more evident for big brands, equipped by large power of spending, able to amortise the ADV costs, favoured by a rich amount of product differentiation.

[7]Quoting Hjorth-Andersen, the insights of reviews concern: "The characteristics (attributes) that a rational consumer would expect from the ordinary use of a given commodity. In a given test (article) in this magazine, a number of variants (brands) is tested on a set of selected characteristics and graded from very *poor* to *excellent*".

In the first years of new millennium, the Google ADV[8] revolution, based on Ad-words strategy[9] and the consequent bootstrapping of On-line markets, created the conditions for a precise estimation of ROI (Return On Investment) due to the marketing expenditure. Since the search product cost became embedded in the mechanism (more ADV investment = more probability to reach the product in less time), the perceived quality benefits the big brand. This asymmetry enhances the discrimination between small brands and the dominant ones (different authors have learned this phenomenon such a case of rich-then-richer tendency, in particular Brynjolfsson[8]). High expenditure sellers, able to minimize manufacturing cost and distribution opportunities, receive higher rankings and more visibility, which improve the chances to show high quality products. Following this new perspective, users participate with more attention to the analysis of quality. Quality became related not only to price attributes, but to several not-price signals, released by platform and available from the users. For example: verbose product description, opinions written directly by users, statistical information about the ranking list, recommendations for similar products, and so forth. Attempts at the complex and cumbersome nature of real price, is revealed by Piccone and Spiegler [41], whose work analyzes how firms can choose a 'format' in which is shown the price, giving consumers hard time for choosing a default firm. This amount of information new issues: helping but building a labyrinth of discorded signals, which prevent clarifying the right relation price-quality.
In his seminal article, Bolton[3] attempts to orient the consumer in this labyrinth, actually comparing trading in a market with online feedback to markets without. The results confirm that the feedback mechanism induces an improvement in transaction efficiency, leveraging benefits typical of public goods, in which the whole community of consumers takes part. Nowadays, the main resources of feedback are the UGC, that the customer can upload to the retailing platform after purchasing a product, distinguished by the technical specification and description put forth by the manufacturer. The more popular and shared UGC are the product reviews, split in two basic signals, a quantitative one, the rating, and a qualitative, the textual information centered around product features. Briefly, the platforms replicate the content product type inherited by the classical "Consumer Reports" magazine, mentioned in the seventies by the literature about product quality. From the nebula of information shown before, the simplest quantitative universal value is the rating, elected by users to become the big product quality proxy.
Different papers published in the 2000s take note of this tendency. As Chevalier and Mayzlin [9] demonstrate, the rating can significantly influence buyers' behavior and have a substantial impact on the success or failure of a product.

[8] Advertising

[9] Google Ads is an online advertising platform launched by Google, where advertisers bid to display textual and image advertisements, service offerings, product listings and others. The winner formula of Ads is simple but very efficient, because it is optimized for a very controlled chain such web services: more marketing investment increased product visibility, more probability purchasing.

Other authors argue that the rating can be used for predicting future sales, taking in argument the product quality: Dellarocas et al.[14]. The next decade heralds a certain skepticism, since the rating system is not blind to assure trustability and reliability of product opinions. Basically, at the beginning, "The Consumer Report" was born as a baseline for a no-profit organization to advocate for the consumer in legislative areas. Whereas the platforms are private corporations, some of them of trillion of dollars. This is an important reason as to why the reviews mechanism and the rating system require further checks. This problem is put in relevance by Lafky[30] who designs an experiment to understand if consumers are motivated to rate by a concern for helping the community of future customers, or just showing gratitude/anger that they feel towards online merchants. In the first case their behaviour could be explained as a signal of an altruism attitude, whereas in the second one, they can obtain some economic incentives, due the hope of getting some agreement with the sellers (it is known that in marketing campaigns, there is a mechanism of reviews paid by sellers). In the experiment the payoffs of raters are manipulated, to observe who helps the buyer or awaits a benefit from the seller. The conclusions explore rating policies, in order to make the rating decision transparent and not imitative.
Some authors, such as Li, Tadelis and Zhou[34] suggest some cases, in which a reputation system controlled by platforms in an automatic and unsupervised way, is able to provide an embedded solution toward the information trustability, in particular for the rating. They evaluate the sellers reputation, generated by consumer feedback due to a new mechanism design. In this design, a Machine Learning algorithm learns how to optimize the UGC signals using a “reward-for-feedback" (RFF) mechanism: the sellers suggest a quality level, the buyers can confirm or not. An anonymous process finds the best match. Rietsche at al.[42] have recently proposed a more generalized checking. For the vast majority of the authors, the relevant factors of feedback, assuring for trustability, function both as the signal of the review and the information related to the reviewer’s profile.
Our research goes in the direction to search trusted signals and to analyze density and accuracy of textual information, in order to adjust the rating to reach a more accurate quality index. Recently, a ML approach to distill and classify reviews information has spread, in conjugation with Natural Language Processing (NLP) tools, enabling the transformation of qualitative signals into quantitative ones. This approach, typical of Computer Science during the last two decades, has started to take roots in economics fields. In different economic research areas have taken advantage from NLP and ML tools, to face a new challenge text information management. Gentzkow et al.[20] have recently proposed a wide survey of this trend. In light of this, we add our contribution in the development area of language engineering.

3 UCG contents and main features

In this section we publicly the structure and the information which it is possible to extract from a UGC, provided on a platform. This stage is necessary because it is a panoramic approach to looking for the useful signals to create the new quality index. By UGC we intend textual (linguistic or quantitative) content that is uploaded by users on an online platform, in the form of free giving [10]. A UGC item is composed by rating, votes and textual content. There are two main agents involved in this process: the reviewer writes the textual content and rates the products, and the voters who evaluate the reviewer's action, through a Thumb Up/Down mechanism. The *rating* is the judgment that a user (origin) gives to another agent of the platform (target: brand, product or other users) about a certain interaction that occurred between them. In our analysis, we take into account ratings to a given product. More trustworthy systems like professional sites are characterized by the presence of an authority who governs the collection and storage of user's ratings. Conversely, several on-line retailers, like Amazon, E-bay and others do not distinguish between certified and amateur contributions. The *textual content* is the description and the opinions of the reviewer about the characteristics of the product. We often refer to the couple rating-textual context as to a *review*. Consumer *votes* represent the Thumb Up/Down mechanism adopted by the consumer to approve or reject a review, shown by the vast majority of on-line retailers at the bottom/top of each review. In particular we consider the ratio between the positive votes and the total votes received by each review. The votes represent helpful feedback to weight the rating of each review, confirming or denying its validity. A reviewer provided by many positive votes enjoys a high *reputation*. We underline that the text and the rating are a double-signal which orients the consumer in its choices about the product, where the vote represents a kind of helpfulness of the pair (rating, text). This is granular information by a large number of consumers (roughly 20 times more than the reviewers). It represents an acceptance or a rejection of the reviewers' opinion.

3.1 Inefficiency of rating

The rating system as main quality signal, as it is designed in the vast majority of platforms, underlies some inefficiencies. We list here the main factors of such inefficiency.

As first factor, we may observe that in many cases *the rating distribution is flat*. The standard platform policies show the average rating of a certain product, using a scale of 9 bars given by the approximation to the .5 digits of such an average (the products are rated according to a 1, 1.5, 2, 2.5, etc. scale). This is not sufficiently detailed to represent a correct signal, as it compresses a lot of different product ratings to the same value, making them indistinguish-

[10] We restrict the domain to a subset of generic UGC contents - images, videos, text, and audio - enough to identify the treatable open source information available for customers.

able. For example, in the consumer electronics market, a consequence of this smooth difference is that top ten brands have an average rating of 4.2 and a standard deviation of 0.02 (statistical research on a sample of 1000 products extracted from Amazon dataset http://jmcauley.ucsd.edu/data/amazon/.)

Second factor, the consumer has to face another issue, namely, an *asymmetry of information* when compared to the platform. In fact, platforms do not show exactly the average rating (due to the aforementioned rounding), and products with similar average ranking are sorted according to a secondary ordering, which is defined by the platform and is completely unknown to the users. Since popular categories have thousands of product, and hundreds with the same average rating, the chosen ordering by the platform necessarily sets the visibility of each product[11].

Third factor is the *uncertainty of the information* provided by the reviews (in particular, by the rating), as a reviewer can (un)intentionally provide *fake reviews*. Specifically, reviewers with material interest in consumers' purchase decisions may post reviews that are designed to push consumers far away from the actual best quality product. For example, during ADV campaigns, it is common practice to acquire reviews from a marketing content providers (like Bazaar Voice[12]). This kind of reviews is called *shill reviews*, i.e., "reviews that distort popularity rankings given that the objective is to improve the online reputation", mentioning Mannino[40][13]. It is not the hard case of review artfully written by humans or machine learning process. Anyway their massive uploading may have deleterious effects on consumer choices. Furthermore, the potential presence of biased reviews may lead consumers to mistrust reviews, as observe Mukherjee[38], Mayzlin[35] and others.

Fourth issue is that *the rating system is incomplete*: acceptable but unremarkable products are not rated because the benefit to the reviewer could be smaller than the cost of providing the rating (in the sense of writing consistent and well motivated reviews). This often happens for new and unknown products. If the consumers themselves believe that reviews of less known products are rare and difficult to find, they may not search for such reviews at all, discouraging the production of new reviews, and decreasing the visibility and the demand of the product. This issue entails some problems relied to the inefficient policies of the online platforms, that tend to privilege big brand, overlooking *niche* and new products fitted in the long tail.

Our fifth and last inefficiency factor is given by a *bias in the information.* For example, comparing the amount of reviews with rating equal to 5 and

---

[11] Amazon adopts a Machine Learning process which exploits the own private information of the reviewer profile, to decide which weight assign, without revealing the details. Following this private criterion, the sort of products ranking is not clear. See https://www.buzzfeednews.com/article/nicolenguyen/amazon-fake-review-problem.

[12] www.bazaarvoice.com

[13] "Shill is a person who writes a review for a product without disclosing the relationship between the seller and review writer. A shill can be the seller or someone compensated by the seller for writing a review. Thus, shills can be agents of sellers, distributors, manufacturers and authors who benefit from the sales of a product".

really short and uninformative textual content (such as "good", "very good", "wonderful"...) with some equivalently small but negative reviews (rating 1 or 2), the first ones dominate significantly the second ones. This is because usually, in a low rating review, the reviewer employs more effort to explain the negative evaluation of the product, when compared to situations in which the reviewer is positive about it. We show some examples of this behavior in Section 4.3: Descriptive statistics. As it is clear, a rating which is not supported by enough information, is not trustable, and this happen more often with positive than with negative ratings, and hence originates the aformentioned bias.

3.2 Reputation

As we have drafted in the related literature, a way to discriminate incorrect reviews is based on the reputation of the agents submitting such information. Several surveys on reputation systems can be found in the literature (Vavilis[46]), in which the main requirements supporting a good or bad reputation are identified. Some examples are focused to discriminate incorrect ratings, in isolate and aggregate way. Other ones draw the needs of sufficient amount of information, in particular about product features: "Features identify the types of trust information that should be considered when assessing reputation as well as the properties of reputation values and aggregation method". The survey shows some limits: the features are seen as theoretical notion, which does not take into account important feasible signals, such as votes, reviews length and other ones. Another constraint is the attention devoted only to peer-to-peer platforms (like E-bay in its old style model, Airbnb, Uber and others), in which both agents needs mutual information during the transaction. Conversely for retailing platforms, the customer is not bounded to activate a communication with sellers, uploading some messages . The possibility to write a reviews is only a voluntary option. In another work, Tadelis[6] points the attention not only on the trustability, but on the quality uncertainty of the message, able to affect the trustworthiness caused by hidden information or hidden actions. In the paper the author sets the conditions under which reputation mechanisms can overcome the problems of asymmetric information. He describes a "trust game", where the key idea is that today's actions will lead to future consequences that affect the prospects of the seller. A best strategy to run a reputation system is to provide future buyers with information about the outcomes of a seller's past behavior. The potential presence of biased reviews may lead consumers to mistrust reviews (see Mukherjee[38], [35]). For this reason distilling the reviewer reputation is necessary for a reliable opinion. Lastly, Michael Luca [32] identifies *tout court* the online consumer reviews, as an adequate substitute of more traditional instances of reputation, finding that consumers react more strongly when a rating contains more additional information, and this information is more visible.

3.3 Votes

Different and heterogeneous approaches to design reputation systems are evident, according to different paths: theoretical studies (Tadelis[6]), add-on tools integrated in the platform (Li, Tadelis and Zhou[34], Rietsche at al.[42]) or laboratory experiments (Lafky[30]). Our proposal exploits the information capturing directly the vote (Thumb Up/Down response on how helpful an online review is) as principal resource. The votes represent an import feature in our model at two levels: the first one is at review level, and is used as rating weight evaluation, the second one at reviewer level, as a proxy of the consensus that he received by a consumer. We discuss how the votes enter in the quality function in detail in Section 5: A quality product index based on UGC).

Different authors sustain the importance of votes such as feedback mechanism. The more interesting contributions come from computer science: see Kim et al.[27], Liu et al.[31], Zhang and Varadarajan [49], Sipos, Ghosh et al.[21]). Following these authors, the rating is submitted to a feedback, exploiting the percentage of positive votes concerning the usefulness of this signal. In economics branches, like Management and Marketing, Korfiatis [28] has been one of the first authors to take into account when the rating is confirmed/deny by the Thumb Up/Down reply. For Mudambi and Schuff[37], "Review helpfulness is defined as the extent to which the customer perceives that a review helps making the right decision in the purchasing process". Other authors who points out its importance are Cao et al.[6]: "Voting and filter options are two systems that are used nowadays to identify helpful reviews".

3.4 Textual content

As we have drafted before, an important component of UGC is the review text, in which the reviewer exposes his impression about product details and conditions of use. One of the simplest attributes of this message put into exam by literature is the length of the textual content, which we will also use in our analysis. Scholars have proved that the length of the review message has a positive impact on review helpfulness (e.g. Liu and Park, 2015; Mudambi and Schuff, 2010; Pan and Zhang, 2011). Long reviews may contain more information (Pan and Zhang, 2011) and more convincing arguments (De Ascaniis and Morasso, 2011) than short reviews. Extensive and well-argued reviews are expected to provide enough information for a consumer to evaluate the quality of the reviewed product (De Ascaniis and Morasso, 2011). Moreover, longer reviews are not only perceived as more helpful in assessing product quality, they are also perceived as more trustworthy than short reviews (Filieri, 2016). Effectively, the length of a review is a measure which is able to show how involved the reviewer is in writing it. From this point of view, consumers may evaluate a reviewer who has spent more time writing a long review to be more credible than one who has written only a couple of lines about a hotel (Pan and Zhang, 2011).

4 Data

In the first part of the section we explain the source of data, the volume and its structure. In the second one the methodology to extract and treat the features ingested as arguments of our quality index. We complete the discussion showing some descriptive statistics about the distribution of the main features in the dataset.

4.1 Dataset

Our dataset consists of information collected on Amazon.com web store within the period 2010-2014, about product price, product characteristics and reviews information of the *consumer electronics category*[14]. Our data have been extracted from a larger dataset, delivered by Julian McAuley, a professor of Computer Science at UCSD University, who has scraped a very large volume of public data from Amazon.com retailing website[15]. The original file about the consumer electronics category contains 1,689,188 reviews related to 10,043 products. We have filtered such dataset in order to capture most popular products and to ensure persistence of such products, i.e. excluding the coming and the going of certain goods. We have taken into account only long presence products, with a life-cycle on the platform longer than 4 years (within the same range 2010-2014), both to ensure a certain homogeneity among comparable goods and because, in shorter periods, the votes are not frequent. This set contains 295 products and 51,264 reviews. Furthermore, we have excluded small reviews, composed by less than 10 words, because in such a case the helpfulness of user contents are less reliable. We have sorted this list in descending order by review length, in order to keep more dense information, cutting a threshold to reach an easily splittable number (100 products), to better separate price segments. The ultimate sample contains 100 products and a total of 29380 reviews.

[14]This category encompasses a massive volume of electronics that includes televisions, cameras, digital cameras, PDAs, calculators, VCRs, DVDs, clocks, audio devices, headphones, tablets, smartphones and many other home products

[15]The DataSet "Amazon product data" (`http://jmcauley.ucsd.edu/data/amazon/`), available for academic use, contains product reviews and metadata scraped from Amazon US (www.amazon.com). The repository includes several information about products: reviews (ratings, text, helpfulness votes), product metadata (descriptions, category information, price, brand, and image features). All the data are available in *json* format, splitted into several hundreds of files (It is not a database, but a repository of information needed to be mutually linked). The full dataset contains over 180 million reviews related to a pool of almost 6 million products including 142.8 million reviews spanning from May 1996 to July 2014. We selected only a subset of the complete information. We observe that the data collection was scraped the 15th of July 2014, all the information are been backdated from this point (that is the daily votes and other information, are picked up through a snapshot in that day, offering a posterior representation of the dynamic process of UGC).

4.2 Feature Extraction

In this section, we show the features list adopted in our framework of product quality evaluation (see Section 5.1: Conceptual Framework). The literature about UGC gets the best features such as significant predictors for the relationship between votes, rating and text properties.

We observe that there is a very small intersection between the UGC literature, that is the field with which we may compare our features extraction and our analysis of the helpfulness evaluation of the rating and the votes, and the economics studies about the more classical quality-price relationship. The works about UGC are often devoted to perform a *self standing* quality information analysis, which is strongly detached not only from products, but from all the concepts of efficiency of the market, asymmetry of the information, and other issues usually of interest from an economics perspective. Our contribution attempts to make the bridge between these two lines of research.

Scholars have investigated the combination effects of numerical and textual characteristics captured from reviews, to reach consumer matching taste, in different contexts. In recent years, sentiment analysis, that is capturing the polarity of the text speech (positive, negative, neutral) has attracted increasing attention from scientific communities and business circles. This attempt of using a text analysis is complementary to the more classical rating evaluation, and tries to improve the accuracy of the current frameworks[9]. To show some examples of useful features to provide qualitative information about the text, we cite the readability index, which is used to reveal how much time the consumer spends for reading a review. In the same direction, a sentiment instance can show the low/high appreciation of goods, the reviews length alerts about the writer effort, and so forth. Another feature whose relevance the literature agrees on is the reviewer reputation (estimated in different ways, such as by using the received votes in all the activities, or across the social interactions with different players). Hong and Hong[25] offer an exhaustive meta classification of the most used features in the literature. The 25 features we used in our framework are a subset of the overall types, collected in two macro-sets: the first ones referred to the reviews and the other ones to the reviewer.

- Review related factors:

  (i) Review depth (Total number of words of a review; review length (sum of chars); review elaborateness)

  (ii) Review readability (the distribution of Part of Speech: Adjectives, Verbs, Names and others, in the text. Other types of metrics)

  (iii) Review age (and product age): number of months in which is visible on platform

  (iv) Rating

  (v) Sentiment Analysis of text. The feature could be seen such as a duplication of rating but it does not work in the same way. The rating is a synthetic impression of the reviewer in a quantitative form, whereas the sentiment is a ML processing of positive and negative polarities of sentences of text. Often the normalized sentiment value is very different from the normalized rating stars.

- Reviewer related factors:
  (i) Reviewer popularity (based on amount of votes received in the overall reviews)
  (ii) Reviewer reputation (based on average of the percentages of positive votes received)

For the complete features list, see Appendix B: Features. We observe that, among the different groups of features, the ones related to textual properties have less relevance (see the evaluation in Section 5.1: Conceptual Framework). This result is compliant to the literature (see the survey Hong and Hong[25] and Rietsche et al.[42]). Anyway a full and detailed comparison about UGC it is not easy, because the samples of the collected reviews are different, and the dimension of the reviews sample is often too small (less than 5000). Conversely, the reviewer reputation results are comparable, because, also by the other authors, are used quantitative signals for evaluation (like votes, rank positions and others), see Hong and Hong [25].

4.3 Descriptive statistics

In this section we present descriptive statistics analysis of the main features. As we can see in Table 1 at page 26, in the upper-side we shown the data of 295 products and 51,264 reviews. Furthermore, we have filtered and removed small and uninformative reviews, in order to collect 100 products (and 29380 reviews), as we show in the lower-side. Note that, in the upper-side of the table, the information are visualized only by rating, however the two sides maintain the same distribution of reviews and reviewer information. All the next statistics are referred to the sample of 29380 reviews aggregated by product. In Table 2 at page 27 the information are splitted in a partition of product segments by price. The results in the tables confirm a direct correlation between rating and the first type of reviewer reputation (quantity of positive votes)[16], but not about the second coefficient (percentage of positive votes) which remains almost invariant in all the lines. That explains the capability of high rated products to attract prolific reviewers (practically all the luxury items), but not necessarily trustable. In the second table, the rating and the price proceed with the same shape.
Another important correlation, in this case moving in the opposite direction. concerns rating and reviews length[17], highlighted in Table 1. This supports our intuition over quality submersed. Although big brand products confirm a stickiness between high rating and high percentage of positive votes, the

[16]A Pearson coefficient $\rho = 0.88$, in the first table and $\rho = 0.94$ in the second one.
[17]A Pearson coefficient $\rho = -0.68$ in the first table.

same trend is observable also in the one or two rated reviews, well written and long. The explanation of this result is confirmed by the reviewer attention about product issues and troubleshooting (often affecting the one rate review), because consumers appreciate useful information about defeats. Since the quantity of one rated reviews is less than one tenth of five rated, because they attract only the writer very specialized about failure, have a distribution more fitted around positive votes. For example, a lot of five rating reviews are short and limited to few words (as 'Num Words' column confirms), and several writers just imitate the behaviour of their colleagues, limiting to spend a quick effort. Whereas the number of words of one rated reviews are more and more informative (the distribution of review length has a peak in the 2-3 rating medium zone, absolutely not in high rating values). Briefly, as we will show in the Section 6.2: Contribution of votes) there are some product low rated candidate to provide more quality, comparing to that one set only by rating, because affected by signal more informative and trustable.
In the second table 1 we shows the same information, trough the products perspective. We have chosen one hundred products (composed by 29380 reviews) in the electronic market, in particular around the camera market category. In our sample, to avoid an imbalanced distribution fitted completely to the accessories market, we have chosen a spectrum partitioned for price values.
low price products (e.g.: cheap accessories, cables), medium-low price (e.g.: compact cameras low-range), medium-high price (e.g.: compact cameras mid-range, camera lens, expensive accessories), high price (luxury segment e.g: compact cameras high-range, Mirrorless, DSLR, particular camera lens). The splitting criterion was that of dividing our dataset in four groups of the same dimension.

1. *low price (e.g.: cheap accessories, cables)* : $p < 30\$$
2. *medium-low price (e.g.: compact cameras low-range)*: $30\$ \leq p < 120\$$
3. *medium-high price (e.g.: compact cameras mid-range, camera lens, expensive accessories)*: $120\$ \leq p < 250\$$
4. *high price (luxury segment e.g: compact cameras high-range, Mirrorless, DSLR, particular camera lens)*: $p \geq 250\$$

Following the partition by price, we note a similar average rating in each segment. The percentage of positive votes confirms the global result of table 1, except for high price. Statistically, the reviews more voted are also they well voted. Furthermore, in the luxury segment, the average rating is highest. Since also the percentage of positive vote are high, all the signals involved confirms high quality. We observe that the density of price distribution green bars of Figure1 is accumulated around low price segment, that reflects the general distribution of Platform electronics market (Coad [10]), in which the number of accessories is 5 times the amount of reference products. In red bars of Figure2 we show the log-price chart of a septile bars price range within (10-1600). We observe that the white spaces around the range values (50-100) is an effect of the sample of less products in this sub-range.

Table 1: Statistics of main features splitted by rating

rating	v%+	reviewer reputation*	reviewer reputation**	lifetime prod.	# words	# reviews
			(all reviews)			
1 Star	0.44(0.40)	0.17(0.33)	2.95(19.89)	63.16(32.74)	105.19(88.43)	1521
2 Stars	0.39(0.43)	0.17(0.33)	2.96(24.08)	70.27(36.45)	106.21(100.11)	2534
3 Stars	0.34(0.43)	0.17(0.30)	3.55(24.95)	72.69(36.39)	95.76(93.25)	12300
4 Stars	0.37(0.46)	0.21(0.32)	4.08(20.43)	70.22(36.33)	96.82(111.78)	14935
5 Stars	0.36(0.47)	0.25(0.32)	4.16(36.03)	71.40(34.26)	78.32(84.43)	19974
All	*0.37(0.40)*	*0.25(0.39)*	*4.33(22.28)*	*51.95(21.44)*	*98.14(94.57)*	*51264*
			(reviews actually used)			
1 Star	0.36(0.47)	0.17(0.30)	3.22(11.29)		66.03(85.22)	884
2 Stars	0.38(0.35)	0.16(0.36)	2.99(13.21)		63.46 (107)	1467
3 Stars	0.33(0.54)	0.16(0.40)	3.92(12.32)		66.34 (57.91)	7025
4 Stars	0.35(0.51)	0.18(0.41)	4.44(9.43)		61.58(64.15)	8634
5 Stars	0.30(0.35)	0.21(0.38)	4.31(10.21)		39.78 (68.27)	11371
All	*0.34(0.35)*	*0.17(0.40)*	*3.77(11.53)*		*59.44 (73.40)*	*29380*

Note: For each column Mean (St. Dev.). The column lifetime product is filled only in the case of reviews used
(*) Average Amount of percentage positive votes achieved by the reviewer for each review
(**) Average of total positive votes achieved by the reviewer for each review

Table 2: Main product feature plus price partitioned in segment by price.

Group type	#items	price ($)	reviewer reputation*	reviewer reputation**	v%+	rating
Low Price	25	16 (7.4)	2.5 (2.4)	0.23 (0.1)	0.43 (0.30)	4.2 (0.7)
Medium-Low Price	25	54 (21)	3.1 (2.0)	0.21 (0.1)	0.40 (0.22)	4.4 (0.5)
Medium-High Price	25	164 (35)	4.3 (3.0)	0.19 (0.2)	0.38 (0.41)	4.6 (0.3)
High Price	25	598 (220)	4.0 (2.7)	0.19 (0.15)	0.60 (0.30)	4.7 (0.1)
All	100	208 (231)	3.47 (0.7)	0.2 (0.15)	0.45 (0.1)	4.47 (0.2)

Note: For each column Mean (St. Dev.) (*) Average of total positive votes achieved by the reviewer for each review (in product segment. (**) Average Amount of percentage positive votes achieved by the reviewer for each review (in product segment)

Figure 1: Price Distribution ($)

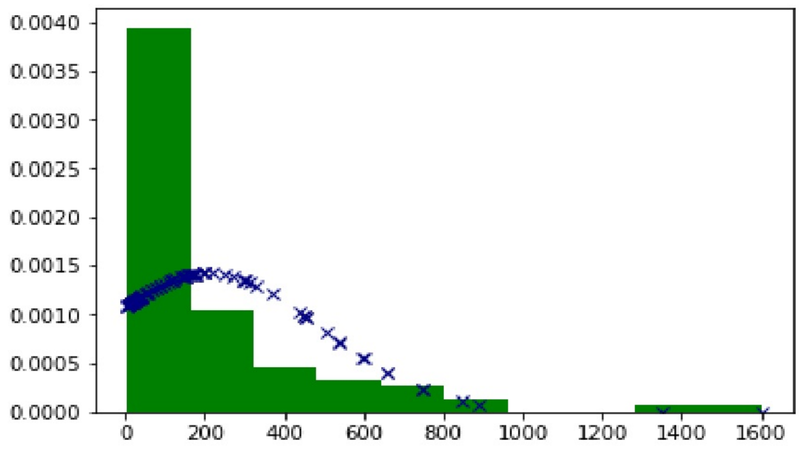
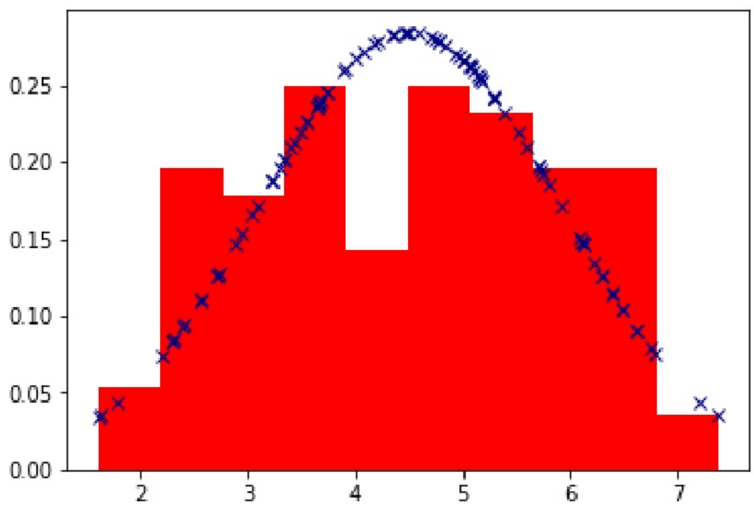


Figure 2: Log Price Distribution ($)



5 A quality product index based on UGC

5.1 Conceptual Framework

After the extraction of the main features (see Section 4.2: Feature Extraction), the scope of our analysis is to select the most salient ones in order to adjust the rating trustability and to define a product quality index for improving the best quality products information achievable by consumers. To avoid falling into risk of confusion between reviews quality, practically concerning information quality, and product quality, we focus the attention about the exact role of the signals. Different authors (Zhang et al. [49], Kim et al.[27], Chua[9]) have confirmed the usefulness based on the inherent properties of the review itself, as a reliable signal: positive in the case of good product or negative in the case of bad one. As discussed before, and visualized in table 6, the rating system is generally accepted by literature as a good quality product index. However, our work puts in relevance the lack in some circumstances (eg.: low price, brand unknown and others), as we have widely illustrated in Section 3.1: Inefficiency of rating. The scope of a new quality product index is trying to trap this inefficiency through some other signals, useful to adjust and stabilize the rating.
The first idea is adopting the vote as the rating weight, because it is a quantitative signal approved by literature as a feedback mechanism of user consensus on the reviewer job (Section 3.3: Votes)[18]. Unfortunately, this value alone is not enough stable to treat uniformly the long/short presence fluctuations of the review on the platform. For example, if a review has taken only two out of two positive votes because the product related is a new arrival, the $\%v+$ will be $= 1$, whereas if another review has 85 positive votes out of 100, the $\%v+ = 0.85$. The less positive percentage is not really a bad effect, because the long-run permanence confirms durable positive feed-back with respect to the other case. To compensate the difference, we are looking for a more robust signal in conjunction with votes, able to dampen these fluctuations. A good candidate is the reviewer reputation rr for different reasons: from a statistical point of view it points out to more relevant reviews, it is also a global propriety of the reviewer and it reinforces the high/low relevance of votes $\%v+$. *A fortiori*, extracting and exploiting only the reviews written by top reviewers (who are ranked in top positions in dedicated platforms pages) seems a good approach. Going back to the previous example, we can learn to vote $\%v+$ as a trustworthiness signal and reputation rr as a magnitude signal. If we multiply $rr * \%v+$[19], for high reputation $rr > 2.5$ this signal amplifies the outcome, reducing the importance of vote in the two cases $\%v+ = 0.85$ and $\%v+ = 1$. When the reputation is low $rr < 1$, the vote acquires more relevance and the

[18]Another candidate could have been the *Brand awareness*, but its effect is an argument investigated in different fields, like laboratory experiments and "Signal Detection" theory, not compliant with our research.

[19]Within all the actual collected reviews, $\%v+ \in [0.5,1]$ and $rr \in [0.5,3.5]$.

difference between the two cases are reduced. In the light of this, our intention is simply to calculate a quality index by linear, or log linear, combination of the arguments: rating, votes and reviewer reputation. Our first goal is deconstructing the information, in order to investigate more formally about the relationship among all the involved variables.

Let us present the basic steps of our process. First of all, we have extracted and collected the most promising features from the UGC resource data, now we must decide which one to choose, evaluating if the reviewer reputation confirms expectations. To get this achievement, we estimate the votes through all the features (see par. *Accuracy evaluation*), in order to extract their contribution for the evaluation. This is an intermediate stage for proceeding toward the feature selection mechanism (see *Features ranking*), verifying the relevance of reviewer reputation. Finally, we need checking for the mutual independence of all the signals through the *Pearson correlation* (see par. *Features independence*)[20], otherwise some components can affect each other, generating a final value distortion. Following this direction, we go to propose a novel formula, as we can see in the equation(3), able to sustain our intuition, on the basis of a statistical learning proof. Briefly, a review provided by high votes and high reviewer reputation is proposed as a good value of reliable information. When the vast majority of the product reviews shows this behavior, and the rating is high, the quality index confirms this outcome, whereas the rating is low, the quality index downgrades the rating. In the case of discordant signals, the result is hybrid and difficult to decode for literature. The power of our quality index is instead the capability to synthesize the quality into a unique measure, both in presence of concordant or discordant signals, to facilitate the comparison between products in all the conditions. We are going to show in detail the main statistical processes to test the consistency of our quality index.

Estimation of the votes

In order to proceed toward feature selection analysis, for finding the weight of the more important features able to affect the vote, it needs to estimate the dependent variable throw a feature regression. The prediction accuracy will permit to decide the more efficient model, from which selecting the meaningful features. In this paragraph we draft the framework of the models' pipeline that we adopt for regression. In the next one we discuss the accuracy of the results.

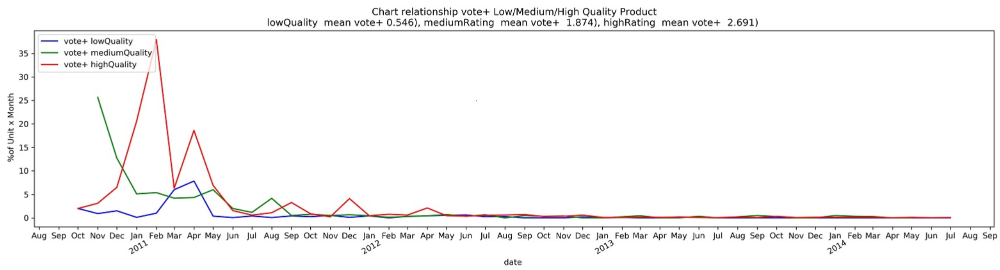
Definition: $\mathbf{v_i}$ is defined as the vector $\mathbf{v_i} = (v_{i1}, \ldots, v_{iT})$, where $v_{i\tau}$ is the *ratio of positive votes for product i at time τ*, i.e., the ratio of positive votes that all the reviews which have been written at time τ (i.e., during month τ) about product i have received:

[20] We observe that, although reviewer reputation is the more weighted argument to predict the vote, the correlation between vote and reviewer reputation is low. The reason is that the regression to estimate the prediction is a complex not linear function, whereas the correlation is a linear test.

$$v_{i\tau} = \frac{\#\text{positive votes for product } i \text{ at time } \tau}{\#\text{total votes for product } i \text{ at time } \tau}, \tag{1}$$

We do not use the total positive votes, because they are distributed in very sparse way with high variance, as we show in Figure3, making us difficult not only the ML estimation, but the complete evaluation of quality index.
The classification setting needs to split the reviews set in a Cross Validation sample: some products will became the Training set, other ones the Test set. Each product i is formed by a set of reviews j, each one composed by a set of properties (features) whose complete list is presented in Appendix B:Features. Given a training set of M samples and a Test set of N ones, in each sample are known the feature review vector $x_1 \dots x_n$. Assuming the vote $v_{ij} = v(x_1 \dots x_n)$, we try to get the best model (regression function) $f(x_1 \dots x_n)_{Train}$ in order to predict the vote $\hat{v} = f(x_1 \dots x_n)_{Test}$ of Test product reviews. Note that the features $x_1 \dots x_n = g_{nlp}(z_1 \dots z_h)$ are a result of $nlp(\cdot)$ function, applied to the original textual features $z_1 \dots z_h$ in order to transform them to quantitative items. We provide a meta-model that launch 7 functions over 25 features and K hyper-parameters (compliant to each model). The regression models used are: Lasso, Ridge, SVR (Support Vector Machine) linear, SVR with kernel rbf (Radial Basis Function), Elastic Net, Decision Tree and Random Forest[21].

Figure 3: absolute vote+ distribution of 3 products during time



Accuracy evaluation

In this task we evaluate the accuracy of the statistical learning model provided by the best prediction. For each models we use a data matrix of 50933 reviews per 25 features (40746 for Training 10187 for Testing)[22]. The best fitting model

[21] All the code used for this paper is available from ("https://github.com/maroccasting/economics-experiments").

[22] The amount of reviews is higher of the effective number used for the index evaluation, but the two quantities do not have the same meaning. In the case of feature evaluation, there are not a privileged aggregation: each unit of information, the review, is treated as independent. In the case of product informativeness evaluation, the reviews are aggregated by product, and the estimation of attributes are referred to each product.

is evaluated with respect to the coefficient of prediction R square[23] R^2 and the Mean Square Error MSE using a Cross Validation partition set (70-30). We find the best fit model in *Random Forest* function.

——————————————— [RandomForest] ———————————————

Total Score: 95.1%

Parameters: 'RandomForest max_depth': 15, 'RandomForest n_estimators': 250[24]

'MSE': '0.01', 'R^2': '0.92', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

It is not easy a detailed comparison with literature, because each experiments use too different samples and different features. In addiction, our scope is not exactly the vote prediction, but the analysis of mutual independence of the features, and the capability to sort them by ranking. We obtained a result of $R^2 = 0.92$, in general very good, taking in account the difficulty of the test (the features are very heterogeneous, and often the votes are accumulated due to mimic effects of emulation. Someone votes a well written review, and she is followed by other people invited by the lighting of this signal. Actually we observes that, when a review gains some votes, the probability to became surrounded by reviews with votes are higher. The Random Forest algorithm[4] provides a feature ranking evaluation, embedded to the model. Through this process it is possible to extract the best features weighted by a Gini Index coefficient[25]. In the last years it is increased the attention of the economists about Random Forest model, to learn typical structures of data text used in economics literature. At this purpose, see some papers of Stefan Wager, Susan Athey et al. [2; 47]. For detailed results of the accuracy of each model, and parameters used, see Appendix C: Accuracy evaluation for each regression model.

[23]The coefficient of determination R^2 of the prediction. The coefficient R^2 is defined as (1 - u/v), where u is the residual sum of squares $((y\_true - y\_pred)^2).sum()$ and v is the total sum of squares $((y\_true - y\_true.mean())^2).sum()$. The best possible score is 1.0.

[24]The hyper-parameters n_estimators means the number of trees, where max_depth means the max number of levels among these trees.

[25]The Gini Index, or Gini impurity, calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly. Optimizing the splitting that better separates the information at the next tree level, it can evaluate the best way of walking through sub-trees of the forest. Let's the intuition: the Gini index varies between values 0 and 1, where 0 expresses the purity of classification (becoming to the same class), and 1 indicates the random distribution across various classes. The value of 0.5 shows an equal distribution of elements over some classes. The direction will be predictor that permits an higher gain during the descend. The same process can be used to ranking the best features, these providing the optimal separability of the paths. The regression task is equivalent to the classification one, from the point of view of the model. It changes only the data output format.

Features ranking

Scope of the feature selection (ranking) process is ordering and reducing the set of features, in order to keep only the more relevant, maintaining enough accuracy level. Through this process we can decide a cut point, in which to take the signal more useful for assuring a good quality index. Practically we want to confirm the relevance of reviewer reputation.
Intuitively, if the original estimation through the full set of features is close to the new estimation through the limited best ones, then can pick up these. More formally: choosing a subset $y_1 \dots y_p$ of original features $x_1 \dots x_n$ so that $\tilde{v} = \tilde{g}(y_1 \dots y_p)$ and $\hat{v} = \hat{f}(x_1 \dots x_n)$, if $|MSE(\hat{f}(\cdot)) - MSE(\tilde{g}(\cdot))| < \epsilon$, where ϵ is the maximum error value admitted, we can adopt the reduced model as a good alternative[26].
The five more ranked features (weight estimated with Gini Index) of Random Forest models, and best setting parameters, are:

Table 3: Features ranking

Feature	**Gini Index**
(MEAN % POS. VOTES per review) per REVIEWER(*)	0.64
NUM REVIEWS OF REVIEWER	0.063
(MEAN TOTAL POS. VOTES per reviews) per REVIEWER(**)	0.037
REVIEW LENGTH	0.022
WORDS QUANTITY (***)	0.017

(*) Average of total positive votes achieved by the reviewer for each review
(**) Average Amount of percentage positive votes achieved by the reviewer for each review
(***) The review length is calculated as number of chars of the full text, whereas the quantity of words is based on words

In the Random Forest regression, as best feature we discover the *reviewer reputation rr*. The normalized Gini weight of reviewer reputation is equal to 0.64, that exceeds roughly ten times the second one[27]. This result is compliant to the literature, as we have shown in Section 3.2: Reputation) Also, the second and third features concerning reviewer reputation, whereas the last

[26]Random Forest (RF) is a statistical learning model, used both for regression and classification task. Intuitively, RF is a sort of generalization of Decision Tree concept. Supposing to walk a tree, in which, for each node, we have to take a decision about which branch to choose. If the decision depends by a condition embedded the node, an algorithm can go across the tree, according to this decision matching. The RF consists in a large number of decision trees like this which operate all together. During the training learning, the algorithm set the conditions withing the nodes of the trees, and the ways to cross them. The model holds a feature selection function, because it knows the decision order in which each feature leads the walking.

[27]As we have shown in a previous note, for RF the feature selection is endogenous to the model. This result is not always true for all models. In general, the feature selection is more difficult and needs a process focused in. A universal approach is the permutation of feature relevance, measuring this value by observing how random re-shuffling (thus preserving the distribution of the variable) of each predictor influences model performance.

twos are referred to the length of the reviews (a sort of density of information: whose meaning is the voter prefers to be well informed). As we will shown in next sections, the reviews length is an impact factor about

——————————— [RandomForest] ———————————

Total Score: 82.81%

Parameters: 'RandomForest max_depth': 15, 'RandomForest n_estimators': 250

'MSE': '0.03', 'R^2': '0.82', 'dim Training Set': $(40746, 5)$, 'dim Test Set': $(10187, 5)$

Since the accuracy of prediction model is very high, we can infer that the vote strongly depends on writer reputation. The index $R^2 = 0.82$ means that, also limiting the training taking only the first 5 best features, we can obtain enough good prediction. Another emerging result is that the sentiment analysis features have low impact to the vote, in other words, the voter is not affected positive or negative information.

Since the reviewer reputation is the more meaningfully decision-maker to split high votes from low ones, we can adopt it as a corrector to amplify the power of votes to adjust the rating, and determinate best product quality. We export this intuition in the quality function equation(3).

Features independence

As we have observed before, we must demonstrate the absence of strong mutual correlation between the features and the rating r involved in the quality index product formula. Not only the main features, but also about all the features $x_1 \dots x_n$ for avoiding dangerous interlacing during the random Forest learning. At this purpose, we calculate the Person correlation between features x_i, x_j. We show only the result about the feature used to weigh the rating, and about the correlation between votes and rating. We show the results only of the main ranked features of table 3 at page 32.

Table 4: Features correlation

Feature	**pearson corr.**
RATING, (MEAN % POS VOTES per review) per REVIEWER(**)	0.12
RATING, NUM REVIEWS OF REVIEWER	0.21
RATING, (MEAN POS VOTES per reviews) per REVIEWER(*)	0.17
RATING, VOTE	0.06
VOTE, (MEAN % POS VOTES per review) per REVIEWER(**)	0.21
VOTE, NUM REVIEWS OF REVIEWER	0.20
VOTE, (MEAN POS VOTES per reviews) per REVIEWER(*)	0.20

(*) Average of total positive votes achieved by the reviewer for each review

(**) Average Amount of percentage positive votes achieved by the reviewer for each review

5.2 Quality index formula

Here, we offer a formal definition of quality. This quality concept will be use in Section 6: Discussion) to analyze its performances. We consider a set of goods $N = \{1, \ldots, n\}$. Each good $i \in N$ is associated to a price-quality pair (p_i, q_i). In our model price and quality are the two pivotal variables leading the choice of a consumer. We do not define the quality of a product as something given and known by the consumer, but as something which she can discover while observing other variables. These variables are observable while investigating the UGC about the product on the marketplace. In particular, we define the *quality of product i* as

$$q_i(r_i, \mathbf{rr_i}, \mathbf{v_i}) = r_i \sum_{k=2}^{T} \frac{V_i(\mathbf{rr_i}, \mathbf{v_i}, k) + V_i(\mathbf{rr_i}, \mathbf{v_i}, k-1)}{2}, \tag{2}$$

where:

(i) $k = 1, \ldots, T$ denotes the discrete unit of time of signal product in the marketplace, where each unit represents one month length period. T is the maximal time horizon that, in our experimental setting, is given by the maximal number of months we consider for a product in the marketplace.

(ii) $V_i(\mathbf{rr_i}, \mathbf{v_i}, k) = \sum_{\tau=1}^{k} log(rr_{i\tau} + 1) \cdot v_{i\tau}$ is a weighted sum, where:

(iia) $\mathbf{v_i}$ is defined as the vector $\mathbf{v_i} = (v_{i1}, \ldots, v_{iT})$, where $v_{i\tau}$ is the *ratio of positive votes for product i at time* τ, i.e., the ratio of positive votes that all the reviews which have been written at time τ (i.e., during month τ) about product i have received (see Section 5.1: Conceptual Framework). Sometimes in tables we denote this signal as $v\%+$ for distinguishing from positive absolute votes, $v+$, when they are present together.

(iib) $\mathbf{rr_i}$ is defined as the vector $\mathbf{rr_i} = (rr_{i1}, \ldots, rr_{iT})$, where $rr_{i\tau}$ is *the component at time* τ *of the reviewers reputation for product* i. We show how these components are calculated. The reviewer reputation is defined through all the reviews she has written on the platform. Given a product i and j a reviewer who has written a review about it on the platform, we define the *punctual reputation* of reviewer j about her review of product i[28], R_{ij}, as the number of absolute positive votes RR_{ij} received by the reviewer about this review, times the percentage of positive vote PP_{ij} received by the reviewer about this review, that is $R_{ij} = RR_{ij} * PP_{ij}$. Given J the set of products about which reviewer j has written a review, we

[28] We suppose, as this is always the case, the each reviewer never writes more than one review about the same product.

define the *reputation* of j as the arithmetic mean of all her punctual reputations:

$$R_j = \frac{\sum_{i \in J} R_{ij}}{|J|}.$$

Then, $rr_{i\tau}$ is defined as the average of the reputations of all the reviewers who have written a review for product i at time τ.

(iii) r_i is the *average rating* that the reviewers have assigned to product i during all the periods of time.

Function V_i represents the cumulative ratio of positive percentage votes the reviews about a given product have received, weighted by the reputation of the involved reviewers, meaning that if most of the reviews are written by "prestigious" reviewers, they receive more weight. On our real data, such a function V_i is convex shaped and lies under the bisector line. The quality is then given as the area under this curve, and it has a higher value if most of the reviews have received mostly positive votes and are written by prestigious reviewers, multiplied by the average rating r_i. Observe that in such a formula, the rating still plays an important role, but it does not represent alone a quantification of the quality of a product.

Note We may observe that the variables $\mathbf{rr_i}$ and $\mathbf{v_i}$ are a quantification of the quality of the UGC about a given product and, as such, they are involved in the formula as monthly values (as they do not represent an absolute value over time, but thet can vary depending on the received reviews). While the rating r_i represents a quantification of a value about the product itself, and as such it may be considered constant in time and it is then used with its aggregated value over the full time period.

6 Discussion

The experimental evidences presented in this section shows that the quality index rating based do not satisfy requirements, able to distill and exploit trustable and useful information spread on platforms. In particular, most rating systems lack support for scope similarity of same quality product with respect to the price, delivering a relationship (price, quality) unable to inform the consumer about correct price clarification. Since, on retailing platforms, most products ordered list and recommended systems suggestions are founded on sorting by rating, this tendency is harmful for the consumer. Because it does nothing but increase the asymmetry between him and the platform, concerning the distribution products informativeness. In particular, the product rating is often flat and opaque, set by platform without exactly being compliant to the average rating of product reviews. In the case of good quality reviews, but not enough to big brand challenging. the rating system contributes to hide the niche and new products, decreasing the products differentiation. Our contribution try to put in relevance this inefficiency, proposing a new solution welfare-driven.

6.1 Products differentiation

The first three tables: table 5,table 6 and table 7 show some attributes of twenty products sorted by rating, belonging respectively to three price segments: low, high and all prices. The segments are a replica of Section 4.3: Descriptive statistics, but in this case we have preferred focused the attention only on low (accessories) and high price (luxury) sectors, because far from each other, and more interesting for the retailing market. These two classes are also compared with overall products. The tables show immediately a small part but representative about the display of products information on platforms: static lists, landing pages, result lists as the results of search engines, recommending system and other cases.
Two points are in evidence: the platform actual rating (represented in first column) is very flat. One implication of this finding is a full arbitrary way to list the results of same rating in the whatever second order, disorienting the consumer when there are thousands of values of 4.5 stars, such happens in electronic consumer market. Furthermore, the variance between the normalized average rating and the normalized quality, favours the last one (0.14 vs 0.06 in low price market products, 0.72 vs 0.07 in high prices). Briefly, in the same samples of 20 items, we find more variety of the prices of products, indexed by quality comparing to these ones indexed by rating. Despite the flatting variance of platform ratings, the average value calculated by algebraic mean and the average of platform rating actually visualized, are pretty similar. The average of qualities is what change. Not only, comparing the ratio between high price and low price rating ($3.97/4.52 = 0.8$), with the ratio between high price and low price quality ($0.72/0.82 = 0.87$), we infer that the quality difference between luxury products and accessories is closer than that shown by the rating differences. The same result can be observed also in table 8, because the average quality of all price products is inferior with respect to the two subsets: low price and high price ones. Furthermore, the best rating products often do not maintain the best quality. Since the platform suggestion, and the consumer attitude, is searching products by rating, the consequences are to choke (putting in secondary pages) high quality products.

Table 5: Low price products

$\mathbf{r}_{pl}$	$\mu\mathbf{r}$	$\mu\mathbf{q}_{norm}$	**price** ($)
4.5	4.50	0.38	11
4.5	4.50	1.00	21
4.5	4.44	0.85	15
4.5	4.39	0.75	11
4.0	4.38	0.95	19
4.0	4.13	0.75	15
4.0	4.10	0.73	25
4.0	4.08	0.90	10
4.0	4.05	0.70	10
4.0	4.00	0.67	18
4.0	4.00	0.85	13
4.0	3.99	0.72	9
4.0	3.98	0.49	27
4.0	3.86	0.72	25
4.0	3.80	0.60	28
3.5	3.58	0.57	10
3.5	3.51	0.60	28
3.5	3.43	0.81	5
3.5	3.40	0.67	10
3.5	3.40	0.71	7
3.90(0.04)	*3.97(0.3)*	*0.72(0.14)*	*15.8(7.3)*

Table 6: High price products

$\mathbf{r}_{pl}$	$\mu\mathbf{r}$	$\mu\mathbf{q}_{norm}$	**price** ($)
5.0	4.94	0.85	1600
5.0	4.90	0.80	305
5.0	4.90	0.85	454
5.0	4.85	1.00	370
5.0	4.79	0.96	540
5.0	4.70	0.74	540
5.0	4.68	0.74	850
4.5	4.64	0.85	440
4.5	4.63	0.83	899
4.5	4.52	0.74	299
4.5	4.45	0.88	314
4.5	4.43	0.82	598
4.5	4.40	0.87	750
4.5	4.40	0.75	454
4.0	4.40	0.87	750
4.0	4.24	0.82	330
4.0	4.09	0.78	450
4.0	4.04	0.69	450
4.0	4.04	0.84	270
4.0	4.00	0.83	620
4.50(0.05)	*4.52(0.4)*	*0.82(0.07)*	*564(298)*

Table 7: All price products

$\mathbf{r}_{pl}$	$\mu\mathbf{r}$	$\mu\mathbf{q}_{norm}$	**price** ($)
5.0	4.90	0.94	454
4.5	4.57	1.00	40
4.5	4.39	0.73	11
4.0	4.22	0.78	25
4.0	4.13	0.74	15
4.0	4.05	0.68	10
4.0	3.99	0.70	9
4.0	3.89	0.80	150
4.0	3.86	0.70	26
4.0	3.80	0.60	28
4.0	3.77	0.70	249
4.0	3.75	0.61	200
3.5	3.66	0.62	175
3.5	3.64	0.59	599
3.5	3.54	0.72	220
3.5	3.53	0.68	200
3.5	3.49	0.60	67
3.5	3.45	0.70	174
3.5	3.44	0.54	170
3.5	3.40	0.55	200
3.85(0.1)	*3.65(0.8)*	*0.65(0.15)*	*151(150)*

Notes: Lines ordered by rating (DEC) Columns 1.st platform rating, 2.mean rating, 3.rd normal quality, 4.þprice. In the bottom line: mean(sd)

A summary is presented in table 8, in which we show some statistical properties (Mean, Standard Deviation) referred to the price distribution within different market segments. Each ones is shown following three types of order (by platform trunk/rounded rating, by average rating and by our quality index). We visualize only the average results of first ten products, which represents the canonical first page consulted by consumers.
This table shows the more meaningful implications of our research. The products price differentiation of all the segments, setting a sort order by our quality

index, is larger then the differentiation achieved by rating order. Actually, for lowest prices (market accessories) the standard deviation of prices is 20% higher ($5.73 > 4.90$). Since this is a relevant segment, in which the number of accessories is five or more times the amount of products, to which are linked, increasing the alternative for the consumer, can change the amount of demand. Just think that *"Amazonbasic"*, the private label brand of Amazon, is composed by accessories for the vast majority of the catalogue[29]. In this segment, the brands are less relevant (eg.: a lot of small and unknown competitors sell headphones or covers for the same mobile phone), then the consumer plans to trust the rating as reference quality index. As consequence of this behavior, the consumer loses two times: she can dispose of less average quality (in table 8 last column, we compare order by quality products, in bold, with order by rating, in italics) and she must renounce to more variety of prices, as shown before.

Table 8: Price Variation of products ordered by different modes

price segments	**pr** mean (sd)	**r** platf mean(sd)	**r** mean(sd)	**qn** mean(sd)
		order by platf. rating (DEC):	first 10 items	
Low	14.50 (4.85)	4.20 (0.02)	4.25 (0.19)	0.78 (0.16)
High	622 (377)	4.90 (0.03)	4.74 (0.13)	0.83 (0.08)
All	75 (135)	4.23 (0.03)	4.18 (0.33)	0.65 (0.20)
		order by average rating (DEC):	first 10 items	
Low	15.50 (*4.90*)	4.20 (0.02)	4.25 (0.19)	*0.77* (0.16)
High	630 (378)	4.95 (0.03)	4.75 (0.13)	*0.83* (0.08)
All	76 (*131*)	4.26 (0.02)	4.18 (0.33)	*0.67* (0.21)
		order by quality index (DEC):	first 10 items	
Low	14.30 (**5.73**)	3.98 (0.03)	4.14 (0.29)	**0.83** (0.09)
High	639 (378)	4.75 (0.02)	4.60 (0.27)	**0.88** (0.53)
All	119 (**140**)	4.20 (0.03)	4.12 (0.38)	**0.78** (0.10)

Notes: Columns 2-3rd price statistics, 4.th platform rating, 5.th mean rating, 6.th normalized quality.

[29]The amount of mobile phones visualized on Amazon.com is around five thousands items, whereas that of headphones is roughly twenty thousands, and the clips and accessories of headphones, that is a subcategory of a subcategory, is one hundred thousand. Even if the average price of these leaf categories is few dollars, multiplied for the large magnitude of this segment, it overpasses the reference product business that drags it.

6.2 Contribution of votes

Table 9: Comparing vote, rating, quality over time among 10 products

years (cod. prod)	μ **v%+**	μ $\hat{\mathbf{v}}$**%+**	**rev** voted (total)	μ **r**	$\mathbf{q}_{norm}$	num words
		Low rating products				
2 (B0044YU60M)	0.66	0.64	121(1279)	3.6	0.56	125
3 (B0054ERGTU)	0.60	0.63	115(1011)	3.3	0.58	110
4 (B004UJ76KD)	0.68	0.65	170(1538)	3.4	0.55	145
5 (**B004GT50YR**)	**0.65**	0.69	301(2110)	3.7	**0.60**	**177**
6 (B0062FR9JK)	0.64	0.68	205(2099)	3.5	0.59	110
mean	*0.65*	*0.66*	*180(1620)*	*3.5*	*0.58*	*136*
		High rating products				
2 (**B0041Q38NU**)	**0.65**	0.68	57(3211)	4.5	**0.60**	**105**
3 (B004G6002M)	0.55	0.51	125(2898)	4.6	0.61	85
4 (B004GF8TIK)	0.61	0.59	210(3285)	4.4	0.67	77
5 (B0043T7FXE)	0.54	0.55	276(3182)	4.7	0.65	60
6 (B0059F9TRQ)	0.55	0.56	311(3300)	4.9	0.64	81
mean	*0.57*	*0.58*	*191(3231)*	*4.5*	*0.64*	*85*

Notes: Compared main attributes of five products low rating (random selected > 1000 reviews) with five products high rating. In the first column the years of life, in the 2nd the % of positive vote, in the 3rd the % of positive vote estimated by our framework, in the 4th the number of product reviews voted and the total, in the 5th the average rating, in the 6th the quality normalized. We have renounced to show products of only 1 year platform presence because affected by too few votes. We do not show products less than 3 stars, because few interesting for the consumer.

The table 9 examines the influence of review length (number of words) and percentage of positive votes, as impact factors with respect to the quality index. More precisely, a bigger quantity of words captures more positive votes, so that to compensate a lower rating. In fact the two products highlighted in bold, whose codes are B004GT50YR and B0041Q38NU, have the same normalized quality index, despite having 1.3 rating difference. We have preferred to choose two groups of random products (both high voted, but one half high rated and the other one half low rated) because a collection of more products, reduced to the average information, are candidate to show flat values. Not useful, to put in relevance the influence of the votes toward the rating, in order to evaluate the quality. In the table, the average difference between high and low rating products is 1.4 (normalizing the rating is roughly 0.3), whereas the difference by quality is 0.8. Which sets our quality estimator as a good candidate for product differentiation, at the contrary of pressed information shown by platform. Which supports our idea of unbalanced rating with respect to the quality. Concerning the vote estimated (hatted), they are close to real values. In some cases, the products' reviews do not have enough votes, especially for new products. In fact the absolute quantity of votes grows up year by year, but in the first months, the lack of votes infers the risk to evaluate an unstable quality. Since we have demonstrate that the vote estimation is not affect by dynamic features whose behavior changes over time, the vote estimation can be used as a proxy. In this way, it is possible to assure a quality index stable and able to assure a good coverage, as we show in the table.

7 Conclusion

The contribution of our research concerns an investigation about inefficiency of rating system to evaluate the product quality. Using Machine Learning frameworks we demonstrate both the (un)trustability of rating, such a signal to easily find a large spectrum of differentiated products, searchable by consumers. In alternative, we propose a novel product quality index based on not-price signals, that combine the rating with other information, publicly available on the platform. In particular the user votes, the reviewer reputation and the review characteristics, able to adjust the rating through a feedback mechanism. Furthermore, our quality index supports a reduction of search costs, and an increase for consumers' chances to search for and compare heterogeneous products. Evidence of experiments, carried out on one hundred products of the consumer electronics market and roughly sixty thousands reviews, confirm the superiority of our quality index to clarify price and quality relationship in new web platforms, exploiting new and traditional features adopted in historical studies. We confirm that, in the range of prices where the information is dense and more uniformly distributed among products, like medium and high price segments in cameras electronic market, the rating of known brands is a good proxy as a quality index, as predicted in most of the literature. However, the different features distribution for accessories, in which the rating is not the clear measure of quality in the vast majority of cases, require the adoption of a more stable quality index.

References

[1] Almagrabi, H., A. Malibari, and J. McNaught. "A survey of quality prediction of product reviews." International Journal of Advanced Computer Science and Applications 6.11 (2015): 49-58.

[2] Athey, Susan, Julie Tibshirani, and Stefan Wager. "Generalized random forests." The Annals of Statistics 47.2 (2019): 1148-1178.

[3] Bolton, Gary E., Elena Katok, and Axel Ockenfels. "How effective are electronic reputation mechanisms? An experimental investigation." Management science 50.11 (2004): 1587-1602.

[4] Breiman, Leo (2001) Random Forests," Machine Learning, 45, pp. 5 - 32.

[5] Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith. "From niches to riches: Anatomy of the long tail." Sloan Management Review 47.4 (2006): 67-71.

[6] Cao, Q., Duan, W. and Gan, Q. (2011), "Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach", Decision Support Systems, Vol. 50 No. 2, pp. 511–521.

[7] Chevalier, Judith and Dina Mayzlin 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." Journal of Marketing Research 43(3): 345-354.

[8] Choi, Hyunyoung, and Hal Varian. 2012. "Predicting the Present with Google Trends." Economic Record 88 (S1): 2–9.

[9] Chua, Alton YK, and Snehasish Banerjee. "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality." Computers in Human Behavior 54 (2016): 547-554.

[10] Coad, Alex. "On the distribution of product price and quality." Journal of Evolutionary Economics 19.4 (2009): 589-604].

[11] Crosby, Philip (1979), "Quality is Free" New York and Scarborough, Ontario: Mentor Books.

[12] Curry, David J. and Peter C. Riesz. (1988), "Prices and Price/Quality Relationships: A Longitudinal Analysis." Journal of Marketing 52, no. 1, pages: 36-51. doi:10.2307/1251684.

[13] Dellarocas, Chrysanthos. "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior." Proceedings of the 2nd ACM conference on Electronic commerce. ACM, 2000.

[14] Dellarocas, Chrysanthos, Neveen Farag Awad and Xiaoquan (Michael) Zhang. 2004. Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning."Planning Proceedings of

the 25nd International Conference on Information Systems (ICIS), Washington, D.C.

[15] de Vries, Lisette, et al. "Explaining consumer brand-related activities on social media: An investigation of the different roles of self-expression and socializing motivations." Computers in Human Behavior 75 (2017): 272-282.

[16] Eliashberg, Jehoshua and Steven Shugan, 1997. “Film Critics: Influencers or Predictors?,” Journal of Marketing, Vol. 61, 68-78.

[17] Ellison, Glenn, and Sara Fisher Ellison. "Search, obfuscation, and price elasticities on the internet." Econometrica 77.2 (2009): 427-452.

[18] Filieri, R. (2016), “What makes an online consumer review trustworthy?”, Annals of Tourism Research, Vol. 58, pp. 46–64.

[19] Fradkin, A., Grewal, E. and Holtz, D. (2018), “The determinants of online review informativeness: Evidence from field experiments on Airbnb”, Ssrn Electronic Journal.

[20] Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. "Text as data." Journal of Economic Literature 57.3 (2019): 535-74.

[21] Sipos, Ruben, Arpita Ghosh, and Thorsten Joachims. "Was this review helpful to you?: it depends! context and voting patterns in online content." Proceedings of the 23rd international conference on World wide web. ACM, 2014.

[22] Ghose, Anindya, and Panagiotis G. Ipeirotis. "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics." IEEE Transactions on Knowledge and Data Engineering 23.10 (2010): 1498-1512.

[23] Hoberg, Gerard, and Gordon Phillips. 2016. “Text-Based Network Industries and Endogenous Product Differentiation.” Journal of Political Economy 124 (5): 1423–65.

[24] Holbrook, Morris B. and Kim P. Corfman (1985) "Quality and Value in the Consumption Experience: Phaedrus Rides Again," in Perceived Quality, Jacob Jacoby and Jerry C. Olson, eds. Lexington, MA: Lexington Books.

[25] Hong, Hong, et al. "Understanding the determinants of online review helpfulness: a meta-analytic investigation." Decision Support Systems 102 (2017): 1-11.

[26] Hjorth-Andersen, Chr. (1984), "The Concept of Quality and the Efficiency of.Markets for Consumer Products," Journal of Consumer Research, 11 (September), 708-18.

[27] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, "Automatically assessing review helpfulness," in the Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, 2006.

[28] Korfiatis, Nikolaos, Elena GarcíA-Bariocanal, and Salvador SáNchez-Alonso. "Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content." Electronic Commerce Research and Applications 11.3 (2012): 205-217.

[29] Kwok, L. and Xie, K.L. (2016), "Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role?", International Journal of Contemporary Hospitality Management, Vol. 28 No. 10, pp. 2156–2177.

[30] Lafky, J. Why do people rate? theory and evidence on online ratings. Games and Economic Behavior, 19(2):554(570. 2014).

[31] Y. Liu, X. Huang, A. An, and X. Yu. Modeling and predicting the helpfulness of online reviews. In Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on, pages 443-452, 2008.

[32] Luca, Michael. "Reviews, reputation, and revenue: The case of Yelp. com." Com (March 15, 2016). Harvard Business School NOM Unit Working Paper 12-016 (2016).

[33] Jin, Ginger, Jungmin Lee, and Michael Luca. "Aggregation of consumer ratings: an application to Yelp.com." Quantitative Marketing and Economics 16.3 (2018): 289-339.

[34] Li, Lingfang Ivy, Steven Tadelis, and Xiaolan Zhou. Buying reputation as a signal of quality: Evidence from an online marketplace. No. w22584. National Bureau of Economic Research, 2016.

[35] Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. "Promotional reviews: An empirical investigation of online review manipulation." American Economic Review 104.8 (2014): 2421-55.

[36] Milgrom, P. and Roberts, J. (1996). Price and advertising signals of product quality. Journal of Political Economy, 4(94):796-821.

[37] Mudambi, Susan M., and David Schuff. "What makes a helpful review? A study of customer reviews on Amazon. com." MIS quarterly 34.1 (2010): 185-200.

[38] Mukherjee, Arjun, Bing Liu, and Natalie Glance. "Spotting fake reviewer groups in consumer reviews." Proceedings of the 21st international conference on World Wide Web. ACM, 2012.

[39] Nelson, P. (1974). Advertising as information. Journal of Political Economy, 4(82):729-754.

[40] Ong Toan, Michael Mannino, and Dawn Gregg. "Linguistic characteristics of shill reviews." Electronic Commerce Research and Applications 13.2 (2014): 69-78.

[41] Piccone, M. and R. Spiegler (2012), 'Price competition under limited comparability', Quarterly Journal of Economics, 127 (1), 97–135.

[42] Rietsche, Roman, et al. "Not all Reviews are Equal-a Literature Review on Online Review Helpfulness." European Conference on Information Systems (ECIS), 2019.

[43] Ruiz, Francisco JR, Susan Athey, and David M. Blei. "Shopper: A probabilistic model of consumer choice with substitutes and complements." Annals of Applied Statistics 14.1 (2020): 1-27.

[44] Scott, Steven L., and Hal R. Varian. 2014. "Predicting the Present with Bayesian Structural Time Series." International Journal of Mathematical Modeling and Numerical Optimisation 5 (1–2): 4–23.

[45] Tadelis, S. (2016). Reputation and feedback systems in online platform markets. Annual Review of Economics, 8:321-340.

[46] Vavilis, Sokratis, Milan Petković, and Nicola Zannone. "A reference model for reputation systems." Decision Support Systems 61 (2014): 147-154.

[47] Wager, Stefan, and Susan Athey. "Estimation and inference of heterogeneous treatment effects using random forests." Journal of the American Statistical Association 113.523 (2018): 1228-1242.

[48] Zhang, Xiaoquan, and Chrysanthos Dellarocas. "The lord of the ratings: Is a movie's fate is influenced by reviews?." ICIS 2006 proceedings (2006): 117.

[49] Z. Zhang and B. Varadarajan, "Utility scoring of product reviews," in the Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA, 2006.

[50] Zhang, Jianqiang, Qingning Cao, and Xiuli He. "Contract and product quality in platform selling." European Journal of Operational Research 272.3 (2019): 928-944.

A Data Collection Details

The format of the files extracted by the data Collection.

- *reviewerID* - ID of the reviewer, e.g. A2SUAM1J3GNN3B
- *asin* - ID of the product, e.g. B01BPFN3S4
- *reviewerName* - name of the reviewer
- *helpful* - helpfulness rating of the review, e.g. 2/3
- *reviewText* - text of the review
- *overall* - rating of the product
- *summary* - summary of the review
- *unixReviewTime* - time of the review (unix time)
- *reviewTime* - time of the review (raw)

The original data contain two of the features, used to calculate the quality of product: the *helpful*, that is the positive consumer votes and all the votes, and the *overall*, that is the rating. About the other feature, the price is obtained through a semi-automatic process, because the price change over time, and we needs the average price. Starting from the *asin* code[30], it is necessary to build the address URL of page product on the platform (e.g.: https://www.amazon.com/gp/product/B01BPFN3S4) then extracting the pricefrom the page. Since the data go back to some years ago, it needs a tool. A simple way is to adopt the site Camelcamelcamel.com, which has the most complete price history of millions of goods sold on Amazon, starting from 2008, and provides a record of Amazon price history over whatever time period. Since the charts of price time series provided by CamelCamelCamel are in jpeg graphics format, it needs another tool to extract quantitative values from the screenshots of price history charts. We have used an algorithm provided by Web-PlotDigitizer (http://arohatgi.info/WebPlotDigitizer/app). In this way we have obtained the time series of prices in digital format, then estimated the average price during time.

[30] ASIN stands for Amazon Standard Identification Number. It is a 10-character alphanumeric unique identifier that's assigned by Amazon.com and its partners. It is used for product-identification

B Features

The full list of features of each review.

1. *DATE* - Date of publication
2. *RATING* - Star rating
3. *RATIO USEFUL / ALL* - Density of useful words[31]
4. *RATIO USEFUL / ALL (First half of the review)*
5. *DENS POS:NVJ* - Lexical feature: (density of main POS Part of Speech: adjective, verb... , ie.: POS: JVN / ALL POS)
6. *DENS POS:NVJ (First half of the review)*
7. *NUM REVIEWS per PROD* - Num reviews of the same product
8. *NUM REVIEWS per REVIEWER* - Reviewer reputation 1: Num reviews written by the reviewer
9. *(MEAN POSITIVE VOTES per reviews) OF REVIEWER* - Reviewer reputation 2: average positive votes achieved by the reviewer for each review (it's independent by the number of reviews)
10. *(MEAN H VOTES per reviews) OF REVIEWER* Reviewer reputation 3: Average Amount of Helpful votes achieved by the reviewer for each review
11. *CURRENT LIFETIME PROD (# months)* Current lifetime cycle of product (until the date of the review)
12. *FULL LIFETIME PROD (# months)* Lifetime cycle of product (until the scraping date)
13. *LENGTH REVIEW* - Length (in chars) of review text
14. *QUANTITY WORDS* - Amount of words of review text
15. *SENTIMENT 1* - Algebraic sum of sentences positive and negative
16. *SENTIMENT 2* - Quantity of positive sentences
17. *SENTIMENT 3* - Quantity of negative sentences
18. *SENTIMENT 4* - Maximum positive contiguous sentences
19. *SENTIMENT 5* - Maximum negative contiguous sentences

[31] For useful words we intends topics useful for the consumer in his choices. The vocabulary of this information has been obtained with a proprietary tool

20. *SENTIMENT 6* - Algebraic sum of sentences positive and negative (First half of the review)
21. *SENTIMENT 7* - Quantity of positive sentences (First half of the review)
22. *SENTIMENT 8* - Quantity of negative sentences (First half of the review)
23. *SENTIMENT 9* - Maximum positive contiguous sentences (First half of the review)
24. *SENTIMENT 10* - Maximum negative contiguous sentences (First half of the review)
25. *ARI Index* - Review readability. This metric is used in linguistic analysis of reviews/news (Readability indicates the extent to which an individual understands and comprehends the product information, which leads to customers accepting information To test the level of understandability of a review, this research examined Automated Readability Index (ARI)

C Accuracy evaluation for each regression model

——————————— [RandomForest] ———————————

Total Score: 95.1%
train_score = 0.98 test_score = 0.95
best_params: 'RandomForest max_depth': 15, 'RandomForest n_estimators': 250[32]
'MSE': '0.01', 'R^2': '0.92', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

——————————— [SVR (rbf)] ———————————

Total Score: 78.83%
best_params= 'rbf C': 100.0, 'rbf cache size': 1000, 'rbf gamma': 0.05, 'rbf max iter': 2000
'MSE': '0.046', 'R^2': '0.725', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

———————————- [SVR_linear] ———————————

Total Score: 73.01%
best_params: 'SVR_linear C': 0.01, 'SVR_linear cache size': 1000, 'SVR_linear max iter': 5000
'MSE': '0.056', 'R^2': '0.662', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

[32] n_estimators means the number of trees, where max_depth means the max number of levels among these trees.

——————————— [Ridge] ———————————

Total Score: 76.83%
best_params: 'Ridge alpha': 0.0001, 'Ridge max iter': 100000.0, 'Ridge normalize': True
'MSE': '0.040', 'R^2': '0.762', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

——————————— [Lasso] ———————————

Total Score: 76.83%
best_params: 'Lasso alpha': 1e-08, 'Lasso max iter': 100000.0, 'Lasso normalize': True
'MSE': '0.040', 'R^2': '0.762', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

——————————- [ElasticNet] ———————————

Total Score: 74.83%
best_params: 'ElasticNet alpha': 2e-05, 'ElasticNet fit intercept': True, 'ElasticNet l1 ratio': 0.7, 'ElasticNet max iter': 100000.0, 'ElasticNet normalize': True
'MSE': '0.046', 'R^2': '0.727', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

Some variants of Random Forest setting.

——————————- [RandomForest SelectFromModel —-
(SGDRegressor(penalty="elasticnet"), threshold='median*0.8') + Original RandomForest]] ———————————

Total Score: 84.75%
best_params: ' max depth': 10, ' n estimators': 250
'MSE': '0.02', 'R^2': '0.88', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

——————————— [RandomForest SelectKBest ——
(score_func=chi2 8 features) + Original RandomForest] ————-

Total Score: 70.78%
best_params: max depth': 10, n estimators': 500
'MSE': '0.04', 'R^2': '0.71', 'dim Training Set': $(40746, 25)$, 'dim Test Set': $(10187, 25)$

Chapter 2

Optimizing Product Quality in Online Search

Abstract

Exploiting an original definition of product quality, based on the information we can get from the User Generated Content, and driven by a statistical learning algorithm, we propose a new ordering mechanism for product search on platforms. This product quality formula is imported in a decision making mechanism which adopts an optimal Stopping Rule, in order to set the optimal time to terminate the search process and choose a good to purchase. We show how the consumer can benefit from the implementation of such a mechanism, demonstrating an improvement in terms of consumer utility at different levels of price, with respect to other sorting traditionally adopted by platforms. We propose a utility function fitted to a Gumbel distribution, and we demonstrate a stochastic dominance of our model. Experimental evidences on the camera market category put in relevance the efficiency of our quality index for ranking the effective quality compared to the more traditional rating system. This is particularly true for the low-price accessory market segment of products, in which we show higher utility dominance and slightly higher elasticity of demand.

1 Introduction

In the first chapter we have proposed a product quality index based on the so-called *User Generated Contents (UGC)*[33], whose insights have been captured by means of Machine Learning tools [5]. The aim of this index is to exploit it to define a mechanism to drive consumers' choices for finding trustworthy signals of actual quality. We have shown how our quality index can improve product price differentiation, applying our model to a sample of electronic camera products. The achieved results are leveraged in this second work, whose main goal is to take advantage of the quality index defined before. Following this perspective, consumer utility is maximized with respect to the search costs. In order to confirm our intuition, we define a theoretical model and we validate on the same sample defined in Chapter 1. This validation leads us to discover that a decision making system, based on our quality index, can improve consumer welfare in products searching. Our proposal is also useful in identifying the sources of search cost heterogeneity and in improving price differentiation of the retrieved products. Our model adheres to a sequential search model,

[33]For UGC we intend any content—text, specially reviews, created and shared by people, publicly available on social media, website, and other marketing channels.

proposing a new implementation, based on a quality index evaluated in an exogenous framework. In this approach, we suggest how to distill the latent information already available on the platforms (ordering best features in statistical learning regression algorithms) to define a quality signal[34]. Such a signal permits us to formalize a decision-making mechanism, called *Stopping Rule* (SR), capable of improving the consumer search. Our model provides two main contributions: *(i)* Firstly, we suggest a new product quality measure that is not only based on easily observable parameters, such as the rating, but also on other indirect signals extracted from the *User Generated Contents* (UGC) *(ii)* Secondly, we make this measure of quality the center of a search mechanism, in order to optimizing consumer choice. Such a mechanism can be implemented as complementary to the classical rankings proposed by the platforms, such as ordering by rating or by price. As a result, we show a better quality ranking. In order to validate our model, we have tested it on data collected by the platform Amazon.com, applying it to one hundred medium and long permanence products on the marketplace (whose information is driven by thirty thousands reviews) belonging to the category *Electronic cameras and accessories.*
In the last two decades, the increased interest of scholars in product information searching has followed the growing quantity of products exposed on platforms. Baye, De los Santos et al.[1] describe the evolution of product search in an analytical survey, from the pre-internet era until nowadays, that shows the importance of search optimization for the stability of competitive markets. Following the first stage (1995-2005) of price-comparison sites, in which the transactions were separated from the search, the platforms established a dominant position over the online retailing. A reason up their incoming supremacy is the integration of both searching and purchasing actions as parts of the same deed. The concomitant introduction of the *Search Engine* tools, spread by the so-called *Google revolution*, changed *de facto* optimization capabilities of the matching between seller-side proposed goods and consumers demand. Unlike village markets, in which the search capabilities were restricted to provide a physical location for product exposure, in the online scenario the visibility is the result of virtual mechanisms. These tools are responsible for searching and ranking, handled by algorithms, as Martens quotes[33]. However, the still imperfect matching between the consumer's needs and the available products is the natural consequence of some frictions. These frictions are related to an asymmetry between the platform, that releases a large amount of information that is incomplete and unclear, and the consumer, who suffers from a restricted and costly time for searching. As a result, the vast majority of the products unexplored and trapped within the *long tail* (Brynjolfsson[8]).

Contrary the popular opinion, the bulk of the information is not always beneficial, as it can mislead the consumer. To overcome this, we simply assume that the search for information is costly, and in order to maximize her

[34] This part is only drafted because treated deeply in the Chapter 1. (See Section 5. quality product index UGC based)

utility, the consumer has to face an important trade-off between the necessity of obtaining good quality information and the need to not waste time. In such an ecosystem, better options for the consumer may be overlooked, if the search time did not give the possibility of discovering the corresponding alternatives. The marketing strategy of most online retailing platforms suffers from some lock-in effects, in which various *best-selling products* keep on first being more salient as firstly highlighted to consumers for a long range of time. Best-sellers increase more and more in visibility, sales and profits. On the contrary, the so-called *niche products* stay stuck in the long tail, reinforcing the well-known law of "the rich get richer and the poor get poorer". This information asymmetry is responsible for a lot of products remained unexplored because trapped and low ranked[8]. In some conditions, these issues are generated by the inefficiency of rating system, when this signal is used as the default measure of all product quality. Some doubts about their capability to cover all scenarios, arise particularly for products not belonging to mainstream brands, such as we have shown in the first chapter (see 3.1 Inefficiency of rating). Conversely, the new quality index proposed here, coupled with a decision mechanism based on the SR, would support consumer search in all these cases. These results are validated by a *Maximum Likelihood Estimation* of the utility function[35] within a large spectrum of reservation prices, starting from accessories and ending luxury products (see Section 6: Estimation). We show that the SR mechanism guarantees the higher utility (stochastically dominant) for the consumer, with respect to the alternative case without SR, under different specifications. In particular, the best benefits of our index are achieved in two segments, accessories (less than 30 $) and medium-high price (100-240$), for different reasons. In the case of the cheapest products, where the brands are often unknown, the reviewers are more interested in writing long realistic and detailed reports. They put more effort into describing product defeats than credits. The medium price level captures the more trustable and verbose reviews too, adding the advantages of known brands. Although the brandization of luxury level group is proficient, in this segment, the probability of finding trustable/trustworthy signals (products that are actually rated highly, sustained by useful information and positive votes) decreases, as does the effective information quality. A lot of reviews are limited to few and unuseful words: "good", "excellent", “wonderful” and so forth.

Concerning the price elasticity of demand[36], the more informative contribution of SR mechanism arises for products whose prices are lower than 100$. In particular, in the accessories market, where the average price is the lowest, our simulation reveals that the “SR” outperforms the alternative without “SR”, showing slightly higher elasticity (-3.5 $< \epsilon_d^{NOSR} < \epsilon_d^{SR} <$ -1). Better

[35]As utility function we use an Extreme Value distribution of Type-I (GEV-I), that is popular in the newest search models.

[36]The own-price elasticity index ϵ_d used here measures the impact of price on search. According to the literature on Searching, we simulate the demand evaluating the difference percentage of products retrieved, in response to a one percent change in price. Supposing also that the product found is also bought.

quality products, captured through our mechanism, explains this result. Conversely, when the range of prices is confined to the luxury segment of the specific category, the demand becomes relatively inelastic (-1 $< \epsilon_d <$ 0), in both cases (with and without SR). Consumers who are willing to remain loyal to brands/products still do so even if the price increases.
The rest of this paper is organized as follows. The next Section reviews the background literature, concerning the Searching about the sequential search model used here, and the SR mechanisms. In Section 3 we present how our detailed data about the reviews are handled through statistical learning algorithms. In Section 4 we define the product quality index, as will be used in Section 5, where we outline the model and the identification strategy. In Section 6 we estimate the model on a data sample of electronic camera category, showing a stochastic dominance of the utility function, between optimal searching with SR mechanism and without. In Section 7 we present the discussion and the conclusion. We terminate proposing possible future direction of our work.

2 Related literature

The recent interest of scholars in search-based demand on platforms has followed the explosion of the online retailing business, and the opportunity for the consumer to view a large amount of products. One of the more investigated problems is how to minimize the effort spent to collect useful signals from the distribution of product information. Though the click-action in online environment can provide a fast and high-performance effect, the spasmodic accumulation of information assumes a trend so high as to counteract the reduction of search time. This imbalance reveals an inefficiency of search tools provided by the platforms, which do not optimize the consumer effort for finding known and unknown products. As pointed out by some researchers, such as Koulayev [32], De los Santos [15] and others, the consumer, looking for unknown products, may estimate the trade-off between expected results that match with his taste and the time spent to achieve them. Their strategy consists in verifying the consumer utility about the products retrieved, trying a few attempts, in order to refine the outcome. Following this perspective, a stable paradigm on which the literature has relied for the last decade, is an extension of the *sequential search model*, which was first proposed by McCall[36]. In this section we will take in account various works that have adopted this model in different situations. Although the economics literature about price search is very broad, the vast majority is focused on questions involving general or partial equilibrium[37], quite different from our research, which is basically devoted to the price-quality relationship. For these reasons, we restricted the comparison

[37] Many models are only theoretical, and take into account simple signals of quality product, far from the data fields. Other works do not enter into details about the search mechanism, but bring out the consumer and firm behaviour for evaluating price dispersion, price obfuscation and similar problems. For a survey see Ellison[18] [27].

to a few essentially close papers. Among them, the studies of Kim, Albuquerque and Bronnenberg[31], De los Santos[15] and Koulayev[32] which have taken into account only the demand side consequences, without modeling adjustments towards the supply side. In chronological order, the first one applies the paradigm of sequential search. Initially, it separates restricted prior searching knowledge, in this case general information acquired ex-ante, in order to proceed toward more detailed evidences (the product pages) discovered step by step. The product searching continues until the marginal benefits overcomes the marginal costs. When they reverse, the search terminates maximizing the expected utility for the consumer. The goal of the paper is to minimize the search cost, for consumers who are well-informed about the product for which they are looking. For this purpose, two ranking list of products are selected: one based on platform sales and another one on recommending list. The model is validated through a dataset of electronics consumer products on Amazon platform (camcorders), similar to the one we used.
The second model, proposed by De los Santos, Hortacsu and Wildebeest [15], uses other kinds of data (books), which allows observation of the online stores visited while consumers shop for an item, and focusing on the store from which the consumer decided to buy. They attack the question in a higher level, quickly to find the store with the best price, comparing "fixed sample size search", which was first provided by Stigler[39], in a seminal paper about economics of information, with the "sequential search" model that we have mentioned before. While, in the case of sequential search, there always exists the possibility of stopping the trial earlier due to sufficiently meaningful results; however, the alternative model requires anticipating a number of alternative n to discover the optimum price among k environments. The authors tackle the research question, focused on price differentiation problem, while considering only the price information and limiting the search to known products. The fixed-search does not estimate a prior sorting in order to facilitate searching, and this conveys the risk of failing at optimizing the search cost when the distribution of price is unfavourable[38]. However, this class of models has contributed the initiation of a rich body of works on retailing research. Nevertheless, in the last decade, the monstrous growth of online information, such as detailed in the introduction, forces this model to set a limited sample ex-ante of well-known products. Currently, the online platforms nowadays submit new product alternatives, offering a complete suite of internal sorting product lists and trapping the consumer inside this world.
The last model, more similar to our research, is proposed by Koulayev[32] by means of a sequential search model. The author simulates the search cost in a popular aggregator of US hotels, capturing all the clicks and pages the consumer visits, until the final transaction is achieved. Different attempts are

[38]Mentioning Baye et al.[3]: "In Fixed sample search, consumers commit to a fixed number, n, of stores to search and then buy at the lowest price at the conclusion of that search. A clear drawback to such a strategy is that it fails to incorporate new information obtained during search, such as an exceptionally low price from an early search".

measured with respect to several setting combinations, obtained by changing the attributes of goods for best performance (eg.: the type of sorting, order by price, by rating, by hotel distances, by filters like panorama view or beach access, and so forth). The consumer's beliefs about these attributes is taken into account by measuring empirical distribution of hotel prices and quality (estimated by not-price attributes), with respect to the distribution of consumer choices driven by the search. In order to optimize the expected consumer utility, the author considers a two-steps function. In the first period, the consumer utilities beliefs are theoretically formulated before new evidence. In a second one, the new information acquired is mapped through to posterior probabilities, using Bayesian inference. Since the posterior can subsequently become a prior again, this inferential chain leads to a result when the system decides the best step stopping. This step is estimated through an advanced regression design based on information extracted from the platform pages and the iterative sequence involving the consumer characteristics vector[39]. Koulayev's model takes into account both product clicking and purchasing, and essentially tries to satisfy the consumer taste with the presence of unknown goods, optimizing his utility. The author adopts a SR mechanism, based on quality-price relationship. Our work is based on SR application as well, but rewards the consumer in a quite different way. Whilst Koulayev employs embedded signals to estimate consumer utilities, our solution treats the utility exploiting the exogenous signal of quality, explained in further detail later. Though less effective in presence of a large amount of products and heterogeneous information distribution, our model is faster than the model used by Koulayev, in the case of small segments of products, and more homogeneous price distribution. In recent years, Morris and Strack [25] have discussed the theoretical common ingredients included in Bayesian sequential models induced by a stopping rule . In light of their work, they have concluded that: "A decision maker sequentially observes signals at a cost and dynamically decides when to stop acquiring information."[25]. The authors' work is paradigmatic and demonstrates the equivalence among different models. These models have in common a cost function used to make a decision through a dynamic evidence accumulation process. The agent decides when to stop the search, because it is no longer worth continuing. In our case as well, the SR mechanism decides to stop when the expected information that would be obtained by continuing is not beneficial. However the step-wise decision is not based on an embedded signal (like the reservation utility of Koulaev's and De los Santos's models), rather on a unique external signal, the previously evaluated quality. The quality formulation has distilled and learned the more useful UGC information, through a statistical learning algorithm, in order to attain the consumer benefit with

[39] The prior-posterior models are studied in Bayesian probability theory. The prior distribution $\mathcal{P}(\Theta)$ represents the belief about the true value of parameters Θ (eg.: in the case of Koulayev[32] consumer characteristics). Whereas the posterior distribution $\mathcal{P}(data|\Theta)$ is the distribution representing our belief about the parameter values after taking the observed data into account (in the example, the data of pages that actually match the consumer characteristics).

respect to the good reached. During the iterative step, when we attempt to find the best quality signal, the consequent utility is a candidate for maximization. From a theoretical point of view, it is not easy to show that the embedded information attained by all the features, in all the paths crossed in the regression methods used in the First Chapter, behaves as a good estimator. The intuition, however, from one side is integrated into a SR theoretical model (Section 5.2), and satisfies the simulation on field data (Section 6).
Concerning the literature related to SR, information search and opinion formation are central aspects of decision making in the choices explained through our model. Here we are facing what is referred to in the literature as an optimal stopping rule problem. Optimal stopping rule problems entail an exploration-exploitation trade-off [34]: a trade-off between exploiting a safe known option, but possibly sub-optimal, and exploring a new unknown one, which may be much better, but may also be much worse, with a consequent loss of time. The most famous optimal stopping problem studied in economics is the *Secretary problem* [19]. It describes a situation in which a firm needs to find the best applicant for a secretarial job. The firm can see each of the applicants sequentially. After seeing an applicant, the firm can rank her quality, and must decide to hire or reject her, without being able to return to a rejected applicant, or predict the quality of future applicants. This differs from other models in that the quality of the applicants is ordinal and not cardinal, meaning that only the ranking matters, instead of the effective value. The optimal solution of this classical problem is extremely elegant and simple, as it consists in interviewing and rejecting the first $1/e(\sim 37\%)$ of the applicants, considering the "value" of the best of these rejected applicants as a threshold, then continuing interviewing and hiring the first applicant who is above this threshold. After that, many other extensions or modifications of the original problem have been investigated, taking into account explicit search costs [40], unknown population size [37], possibility of recall [22], and many other features.

3 Data and Product Features

3.1 Data Collection and Description

The data used in our analysis refer to the so called "*search goods*" category[40]. We use the same dataset of *consumer electronics camera* information we have presented in Chapter 1 (Section 4). This choice is explained by the fact this is one of the fastest-growing e-commerce categories, and studying the product information acquisition of durable search goods, like in this category, is more feasible for estimating demand primitives (the consumer can explain its choices, without deep experience and evaluation after purchasing the good). Our dataset consists of information collected on Amazon.com web store within the period 2010-2014, about product price, product characteristics and reviews information. We have filtered the full dataset to reach products provided with more information compliant to the analysis requirements. At the end of the process, our dataset counts 100 products and 29380 reviews, from which we have extracted 25 product features.

3.2 UGC information

In the first chapter, we have illustrated the insights of UGC, and why they are useful to build a trustable product quality signal. Here, we do not repeat all the properties and parameters setting in order to define which UGC are the most appropriate to define our index. However, We present a brief summary about the meaning of User Generated Content (UGC), and we refer the reader to Chapter 1 (Section 3) for more details. For UGC we intend a generic information (linguistic or quantitative) that is uploaded by some users on an online platform. A UGC item is composed by rating, textual content and votes. The *rating* is the judgment that a reviewer gives to a certain product, in the form of stars number (generally from 1 to 5), the *textual content* is the description and the opinions of the reviewer about the characteristics of the product. The *consumer votes* represent the Thumb Up/Down mechanism adopted by the consumer to approve or reject a review. A reviewer provided by many positive votes enjoys a high *reputation*.

[40]The literature differentiate between two product types a) *search goods* and b) *experience goods*. Search goods can be generally described as products of which the characteristics can quite easily be evaluated before purchasing them (Luan et al., 2016). Most products (and in particular, the ones of our analysis) fall into the search good category, such as clothing, home furnishings, electronic objects, etc. Experience goods can be generally described as products of which the characteristics are quite difficult to evaluate prior to buying them (Baek et al., 2013). It is very important to differentiate between the two types of products because the product type influences the effect of some factors on the helpfulness. consumers may, for example, rely on prior experience, on product inspections or in some other ways for searching and collecting information, such as recommendation or word of mouth. In the most recent scenario and as it is the case of our analysis, such a search of information may be done online thanks to the so called UGC.

3.3 Price segments

Our dataset has been splitted in four price segments, for different reasons. First of all, both the reviewer and consumer attitudes change at the variation of price, in term of attention, precision and requirements. This heterogeneity of signals, basically changes the quality distribution, as we have already shown in the Section 6 of Chapter 1. The price segments listed here concerns: low price products, medium-low price, medium-high price, high price. The splitting criterion was that of dividing our dataset in four groups of the same dimension.

1. *low price (e.g.: cheap accessories, cables)* : $p < 30\$$
2. *medium-low price (e.g.: compact cameras low-range)*: $30\$ \leq p < 120\$$
3. *medium-high price (e.g.: compact cameras mid-range, camera lens, expensive accessories)*: $120\$ \leq p < 250\$$
4. *high price (luxury segment e.g: compact cameras high-range, Mirrorless, DSLR, particular camera lens)*: $p \geq 250\$$

4 Product Quality Index

Our work proposes a new quality index based on UGC. From one side, it inherits the current perspective to adopt the rating as basic signal for quality. From the other side, it adjusts it through a feedback mechanism. This feedback is exploited through several key features, extracted from public available online information. The extraction of the key features is presented in details in Section 5 of Chapter 1. In the following, we present a short overview of how this extraction has been implemented.

4.1 Statistical learning algorithms

The rating system is generally accepted by literature as a good quality product index. However, our work puts in relevance the lack in some circumstances (e.g.: low price, brand unknown and others), as we have widely illustrated in the first chapter Section 3.1 Inefficiency of rating. The scope of a new quality product index is trying to trap this inefficiency through some other signals, useful to adjust and stabilize the rating. An approach could be adopting the vote as the rating weight, because it is a quantitative signal approved by literature as a feedback mechanism of user consensus on the reviewer job. Unfortunately, this value alone is not enough stable to treat uniformly the long/short presence fluctuations of the review on the platform, and needs a stabilizer mechanism able to dampen these fluctuations. A good candidate is the reviewer reputation rr: from a statistical point of view it points out to more relevant reviews, it is also a global propriety of the reviewer and it reinforces the high/low relevance of votes $\%v+$. In the light of this, our intention is simply to calculate a quality index by linear, or log linear, combination of

the arguments: rating, votes and reviewer reputation. Our first goal is deconstructing the information, in order to investigate more formally about the relationship among all the involved variables.
In this chapter we draft only the basic steps of our process, referring to the first chapter Section 5.1 Conceptual Framework, the detail of this operation (all the paragraphs mentioned are referred to first chapter Section 5.1). First of all, we have extracted and collected the most promising features from the UGC resource data, now we must decide which one to choose, evaluating if the reviewer reputation confirms expectations. To get this achievement, we estimate the votes through all the features (see parag. *Accuracy evaluation*), in order to extract their contribution for the evaluation. This is an intermediate stage for proceeding toward the feature selection mechanism (see parag. *Features ranking*), verifying the relevance of reviewer reputation. Finally, we need checking for the mutual independence of all the signals through the *Pearson correlation* (see parag. *Features independence*). All these signals concur toward an implementation of *quality*. Following this process we propose a novel formula, in the equation (3).

4.2 Quality formula

Here, we present the formal definition of quality we have already extensively described in Section 5.2 of Chapter 1. We present again the formula in order to facilitate the reading of the current section.

We consider a set of goods $N = \{1, \dots, n\}$. Each good $i \in N$ is associated to a price-quality pair (p_i, q_i). In our model price and quality are the two pivotal variables leading the choice of a consumer.We do not define the quality of a product as something given and known by the consumer, but as something which she can discover while observing other variables. These variables are observable while investigating the UGC about the product on the marketplace. In particular, we define the *quality of product i* as

$$q_i(r_i, \mathbf{rr_i}, \mathbf{v_i}) = r_i \sum_{k=2}^{T} \frac{V_i(\mathbf{rr_i}, \mathbf{v_i}, k) + V_i(\mathbf{rr_i}, \mathbf{v_i}, k-1)}{2}, \tag{3}$$

where:

(i) $k = 1, \dots, T$ denotes the discrete unit of time of signal product in the marketplace, where each unit represents one month length period. T is the maximal time horizon that, in our experimental setting, is given by the maximal number of months we consider for a product in the marketplace.

(ii) $V_i(\mathbf{rr_i}, \mathbf{v_i}, k) = \sum_{\tau=1}^{k} log(rr_{i\tau} + 1) \cdot v_{i\tau}$ is a weighted sum, where:

(iia) $\mathbf{v_i}$ is defined as the vector $\mathbf{v_i} = (v_{i1}, \dots, v_{iT})$, where $v_{i\tau}$ is the *ratio of positive votes for product i at time τ*, i.e., the ratio of positive votes that all the reviews which have been written at time τ (i.e., during month τ) about product i have received:

$$v_{i\tau} = \frac{\#\text{positive votes for product } i \text{ at time } \tau}{\#\text{total votes for product } i \text{ at time } \tau}, \tag{4}$$

(iib) $\mathbf{rr_i}$ is defined as the vector $\mathbf{rr_i} = (rr_{i1}, \ldots, rr_{iT})$, where $rr_{i\tau}$ is *the component at time* τ *of the reviewers reputation for product* i. We show how these components are calculated. The reviewer reputation is defined through all the reviews she has written on the platform. Given a product i and j a reviewer who has written a review about it on the platform, we define the *punctual reputation* of reviewer j about her review of product i[41], R_{ij}, as the number of absolute positive votes RR_{ij} received by the reviewer about this review, times the percentage of positive vote PP_{ij} received by the reviewer about this review, that is $R_{ij} = RR_{ij} * PP_{ij}$. Given J the set of products about which reviewer j has written a review, we define the *reputation* of j as the arithmetic mean of all her punctual reputations:

$$R_j = \frac{\sum_{i \in J} R_{ij}}{|J|}.$$

Then, $rr_{i\tau}$ is defined as the average of the reputations of all the reviewers who have written a review for product i at time τ.

(iii) r_i is the *average rating* that the reviewers have assigned to product i during all the periods of time.

4.3 Asymmetry in quality-rating relationship

Rifling through product reviews of amazon.com website, we have observed that the information distribution is not symmetric. Especially, going across the high-low rating axis, the richer product reviews are not always the ones with the highest ratings. On the contrary, a lot of one-or-two-rated reviews provide useful and eloquent information about the product. In this section, we investigate this phenomenon, showing the relationships between price, rating and quality, within different conditions. This investigation occurs as a preliminary analysis for comparing price regions in which quality and rating are directly proportional to each other (and then, in which the rating could work as a good approximation for quality), to other regions in which they have opposite behaviors. In fact, as we will show by validating equations (5) on our data, the probability to find high quality products is higher when both price and rating are high. On the contrary, this is not true anymore in the case of

[41] We suppose, as this is always the case, the each reviewer never writes more than one review about the same product.

low prices. The first result is conventionally accepted in the retailing literature, for village markets and online platforms. But the second one is often not taken for granted, and it has been largely overlooked by researchers.

Given the 100 products of the sample we described in Section 3.1, we split it in two according to the rating: "Low Rating" and "High Rating" products, as separated by the median value, which is equal to 4.2.

1. Low Rating (rating $r_l \leq 4.2$)
2. High Rating (rating $r_h > 4.2$)

Another partition for our dataset is the one which separates our sample between "low price" and "high price" products, as divided by the median value, which is equal to 100\$[42].

1. Low Price (price $p_l \leq 100\$$)
2. High Price (price $p_h > 100\$$)

Finally, a third partition is about quality: "Low quality" and "High quality" products, as divided by the median value, which is equal to 0.7.

1. Low quality ($q_l < 0.7$)
2. High quality ($q_h \geq 0.7$)

Statistical evidence of asymmetry. Given an attribute $a \in A = \{r_l, r_h, p_l, p_h, q_l, q_h\}$ (where r_l is for low rating and r_h for high rating, etc.), we consider the probability $\mathcal{P}(a)$ that a random product taken from the dataset belongs to the given group. This probability is defined as the number of products in the corresponding group divided by the total number of products:

$$\mathcal{P}(a) = \frac{\#\text{products tagged by attribute 'a'}}{\#\text{ all products}}.$$

For example, if we refer to high quality products, then $a = q_h$ and we have

$$\mathcal{P}(q_h) = \frac{\#\text{ high quality products}}{\#\text{ all products}}.$$

Given $a_1, a_2 \in A$, the conditional probability is defined as

$$\mathcal{P}(a_1|a_2) = \frac{\mathcal{P}(a_1 \cap a_2)}{\mathcal{P}(a_2)} = \frac{\#\text{ products tagged by both attributes 'a1' and 'a2'}}{\#\text{ products tagged only by 'a2'}}.$$

[42] Observe that in this case we have chosen a different partition with respect to prices compared to the one we have shown in Section 3.3, since a tetra-partition is not appropriate for our current analysis, as it would lead to complex and not very useful cases.

For example,

$$\mathcal{P}(q_h|r_h) = \frac{\mathcal{P}(q_h \cap r_h)}{\mathcal{P}(r_h)} = \frac{\text{\# high quality AND high rating products}}{\text{\# high rating products}}.$$

We illustrate our conjecture by presenting the following inequalities

$$\begin{cases} 1)\ \mathcal{P}(q_h|p_l \cap r_h) < \mathcal{P}(q_h|r_h) < \mathcal{P}(q_h|p_h \cap r_h) \\ 2)\ \mathcal{P}(q_h|p_l \cap r_l) > \mathcal{P}(q_h|r_l) > \mathcal{P}(q_h|p_h \cap r_l) \end{cases} \tag{5}$$

that we want to validate on our dataset. Obviously this is not enough to verify the consistency of our quality index and as a full support of our conjecture. However, this analysis is useful to draw the intuition and to provide a qualitative preliminary analysis of how our features behave.

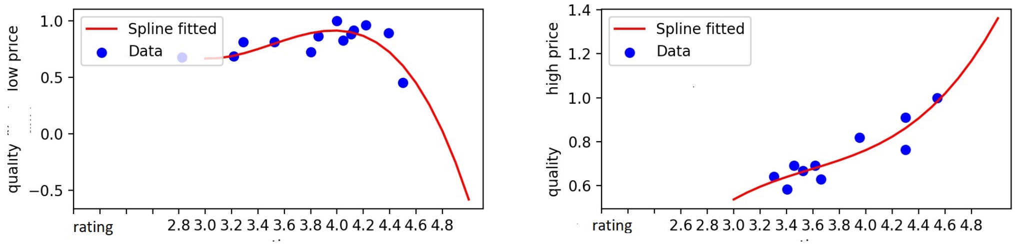
Applying the Bayesian rule, the conditional probability of the first expression $\mathcal{P}(q_h|p_l \cap r_h)$ can be written as $\frac{\mathcal{P}(q_h \cap p_l \cap r_h)}{\mathcal{P}(p_h \cap r_h)}$. We have that $\mathcal{P}(q_h \cap p_l \cap r_h) = \frac{\#\{i|q(i) \geq 0.7 \wedge p(i) < 100\$ \wedge r(i) \geq 4.2\}}{\#\{\text{total } i\}} = \frac{2}{30} = 0.06$ and $\mathcal{P}(p_h \cap r_h) = \frac{\#\{i|p(i) \geq 100\$ \wedge r(i) \geq 4.2\}}{\#\{\text{total } i\}} = \frac{13}{30} = 0.43$, from which we get that $\mathcal{P}(q_h|p_l \cap r_h) = \frac{0.06}{0.43} = 0.13$. In a similar way, we calculate the values for the other probabilities, from which we obtain:

$$\begin{cases} \mathcal{P}(q_h|p_l \cap r_h) = 0.13 < \mathcal{P}(q_h|r_h) = 0.22 < \mathcal{P}(q_h|p_h \cap r_h) = 0.67 \\ \mathcal{P}(q_h|p_l \cap r_l) = 0.15 > \mathcal{P}(q_h|r_l) = 0.12 > \mathcal{P}(q_h|p_h \cap r_l) = 0.09 \end{cases}$$

We observe that the difference between the extreme values of the first inequality chain in (5) that is equal to $\mathcal{P}(q_h|p_h \cap r_h) - \mathcal{P}(q_h|p_l \cap r_h) = 0.67 - 0.13 = 0.54$. This is bigger then the same in the second inequality chain, which is equal to $\mathcal{P}(q_h|p_l \cap r_l) - \mathcal{P}(q_h|p_h \cap r_l) = 0.15 - 0.09 = 0.06$. We can conclude that the more relevant asymmetry happens between high prices with respect to low ones, when considering high rating products. Instead, such asymmetry is irrelevant when considering low rating products.

We summarize these results, observing how the probability of finding cheap products, when both the rating and the quality are high, is pretty low. Since the canonical quality index of platforms is the rating, when the user sorts the products by price and after by rating, she will get a quality dispersion in the winding of the long tail. Hence, in the case of accessories (i.e., in the case of low price products), a product list ordered by rating does not inform the consumer about the correct quality. On the contrary, over a certain price threshold, the rating is more in line with the quality. A possible explanation may be found in the consumer behavior. In fact, usually a consumer puts more attention in reading a review, when she does not recognize the brand (and the accessories are more often unbranded). On the contrary, in the case of high price products and mainstream brand awareness, she accepts more easily the rating assigned by the reviewer without further investigation.

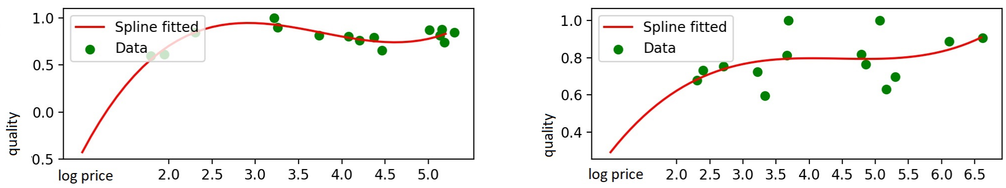
Graphical representation of asymmetries. In Figure 4, we show the asymmetries we observed in the previous inequality through a chart representation of the quality as a function of the rating, i.e., representing the $q(r)$, in the two previously mentioned segments of price, i.e., low and high. The first function, for the low price group, is given by a cubic spline approximation of the fitted sample, and it is concave and decreasing. Conversely, the second function is convex and increasing. The two slopes confirm the inequalities we showed before.



((a)) low price level ((b)) high price level

Figure 4: Comparison of low-high price segments (x-axes: rating, y-axes: quality)

Similarly, in Figure 5 we represent $q(log(p))$, i.e., the quality as a function of the logarithm of the price, for the low and the high rating groups. The two curves have a very similar shape. Briefly, in term of interpolation fitting, changing the rating level, the distribution of data affects less the quality.



((a)) low rating level ((b)) high rating level

Figure 5: Comparison low-high rating level (x-axes: log price, y-axes: quality)

5 Theoretical model

As we have already mentioned in the introduction, the imperfect matching between the needs of the consumer and the available products on online markets is the consequence of some information asymmetries. Such asymmetries were already existing previously to the Google revolution, but became surprisingly more evident in the online ecosystem. In fact, from the one side, the uncontrolled amount of UGC gives, in theory, the potential to the consumer to collect a more detailed information. On the other side, the way in which the platforms partially show and partially hide this information, and the limited attention of the consumer for getting and organizing this giant flow of news, make the online world more complicated and more inefficient than expected. In order to reduce this asymmetry, the consumer needs to get as more good information as possible out of these UGC, but by doing that, she faces a very important trade-off, between the wish of discovering the best quality product, and the search costs among different alternative (see Section 5: Theoretical model), which are becoming higher and higher, due to the required effort for doing that.

We represent this trade-off as an optimal stopping rule problem, in which a consumer makes a decision about the optimal time for stopping searching new alternatives, and start exploiting one of the known ones. She has to decide when to be enough satisfied with a sub-optimal choice, given the cost of investigation, better than keep on looking for a better alternative. The potential of this optimal stopping rule approach, compared to the more classical optimization problems in economics, is twofold: *(i)* firstly, the decision of stopping does not need to be taken *a priori*, but can evolve with the investigation of the scenario by the consumer. This feature translates well the fact that, while looking for a good alternative, the time of investigation often depends on and adapts to the past history of the investigation itself: an investigation may be extremely fast when, by lack, a very good option is soon discovered , while it can be extremely long in the opposite case. *(ii)* Secondly, this *a posteriori* decision does not need for the consumer to have a preliminary guess about the distribution of the quality, or any Bayesian updating of some beliefs about it. The result of this stopping rule problem is then embedded into a 2-threshold model, for fixing the minimum level of quality a consumer should aim at getting while pursuing a good. The other threshold is represented by a maximum price the consumer is willing to pay and that represents a budget constraint. This threshold is assumed to be given a priori and it is not further discussed in our analysis.

5.1 The 2-threshold (price-quality) decision model

Given the set of goods N and the couple price-quality (p_i, q_i) for each $i \in N$, where p_i is given and q_i is observable through equation (3), the consumer checks each good one after the other, according to a given ordering $\pi = j_1, \ldots, j_n$, with the meaning that if $i = j_t$, then i is the t-th good which has been checked by the consumer. We suppose that the consumer makes a choice about one single unit of good to pursue[43], according to the following 2-threshold model in which she fixes:

1. a *maximum price threshold* $\bar{p}$ she is ready to spend for pursuing the good;
2. a *minimum quality threshold* $\bar{q}_t$, denoting the quality a consumer requires from a good, and that *a priori* may depend on the time t of investigation.

In our 2-threshold model, we suppose that the consumer pursues a good as soon as both the conditions are met. We assume that the price constrain remains the same, while the wished quality may adapt with time, depending on the

[43]Differently from some more classical models, in which the chosen quantity to pursue is the main variable, in these models we assume that the choice is about which good to pursue, and the quantity is always equal to 1.

collected information. This assumption translates the fact that a consumer has immediately all the information about the prices, without effort, and then she is capable of fixing a threshold a priori, while establishing the quality requires some cost of search and, consequently, also the level of satisfaction may adapt with time. According to our 2-threshold model, given a product $i \in N$, and a couple price-quality (p_i, q_i), a *necessary* condition for the consumer to pursue good i at time t is that

$$\begin{cases} p_i \leq \bar{p} \\ q_i \geq \bar{q}_t, \end{cases}$$

with $i \in \{j_1, \ldots, j_t\}$, meaning that it is necessary for the consumer to have already discovered good i at time t in order for her to decide to pursue it. Such a condition is not sufficient, as the consumer could have already chosen another good at an earlier time than t. The goal of establishing the optimal satisfying threshold $\bar{q}_t$ for the quality at time t can be represented as a search problem, that we model and investigate in the following section.

5.2 The optimal stopping rule problem

In a classical search problem, a player has to face an important trade-off, between the wish to look for a best alternative, and the limited amount. This kind of problems are usually modeled as *stopping rules problems*. Classical stopping rules problems are usually defined by:

1. a sequence of random variables $X_1, X_2, \ldots$ whose distribution is assumed to be known, and
2. a sequence of real-valued reward functions $u_0, u_1(x_1), u_2(x_1, x_2), \ldots$.

A player may observe the sequence $X_1, X_2, \ldots$ for as long as she wishes. For each $t = 1, 2, \ldots$ after observing the signals $X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t$, she may stop and receive the known reward $u_t(x_1, \ldots, x_t)$, or she may continue and observe the signal X_{t+1}. In the literature, the solutions to such problems are often well representative of the real life experience, in which being satisfied with a sub-optimal choice may be better than wasting too much time for investigation.

This kind of model matches well with our online ecosystem, in which the enormous amount of information makes really difficult, for a consumer, to have the time to fully explore all the alternatives while looking for the best one and in which getting the necessary information to establish the quality of a good is assumed to be costly.

In our model the player is a consumer, and the signal (the quality) about a given choice $i \in N$ (a product, a search good) may be observed by spending enough time checking all the information about a given product, namely, checking the UCGs about it available on the platform. We consider the maximum number of steps for investigation, i.e., the time horizon, equal to n, i.e., equal to the total number of goods. In the worst case, in fact, the consumer may decide to check all the goods before making her choice. We suppose tha investigating one of the alternatives has a fixed cost per time unit equal to c. Then, for each $t = 1, 2, \ldots, n$, we define the sequence of real-valued reward functions for the consumer as

$$\begin{aligned} u_0 &= 0 \\ u_t(q_{j_1}, \ldots, q_{j_t}) &= max\{q_{j_1}, \ldots, q_{j_t}\} - ct. \end{aligned} \tag{6}$$

This is the reward function of a classical stopping rule problem, well known as the *job research problem*, with recall, when the exploration of an alternative has a cost and when the horizon of time is finite. We can observe that, when $c = 0$, i.e., when investigating for the quality of a good is not costly, the best strategy for a consumer is to investigate all the available alternatives, before making a choice. The function $u_t(q_{j_1}, \ldots, q_{j_t})$, described in (6), represents the utility function for a consumer when pursuing the best $\arg\max\{q_{j_1}, \ldots, q_{j_t}\}$ at time t of exploration.

With Theorem 1, we obtain the quality threshold to maximize this utility, i.e., the optimal stopping rule for the problem defined in (6), when we suppose that the quality is uniformly and independently distributed.

Theorem 1. *Suppose, without loss of generality, that the ordering is given by $\pi = 1, \ldots, n$ and that the corresponding quality is given by the realizations of some random variables Q_i which are uniformly and independently distributed in $\{1, \ldots, Q\}$. Then, the optimal stopping rule problem described in* (6) *is given by a constant threshold equal to*

$$\bar{q}_t = \bar{q} = \frac{2Q + 1 - \sqrt{8Qc + 1}}{2}. \tag{7}$$

Proof. Given the finite number of steps, the optimal stopping rule for the consumer can be obtained by *backward induction* in the following way.

Step n-1: We suppose that the consumer has already checked the first $n-1$ alternatives. Let M_{n-1} be the random variable maximum of the first $n-1$ observations. The agent knows its realization m_{n-1}, as she knows the realization q_i of each random variable Q_i with $i \in \{1, \ldots, n-1\}$, results of all the past observations. If she decides to stop and not to investigate the last alternative, she can get a certain utility equal to $m_{n-1} - (n-1)c$, while if she decides to continue and explore the last alternative, she gets an uncertain utility, whose expected value is equal to

$$\begin{aligned} &\mathbb{E}(M_n | M_{n-1} = m_{n-1}) - nc \\ &= \frac{m_{n-1}}{Q} m_{n-1} + \frac{1}{Q} \sum_{j=m_{n-1}+1}^{Q} j - nc. \end{aligned}$$

Then, at step $n-1$ it is convenient to check the last alternative if

$$
\begin{aligned}
&\frac{m_{n-1}^2}{Q} + \frac{Q(Q+1)}{2Q} - \frac{m_{n-1}(m_{n-1}+1)}{2Q} - m_{n-1} > c \\
&\Leftrightarrow m_{n-1}^2 - (2Q+1)m_{n-1} + Q(Q+1) - 2Qc > 0 \\
&\Leftrightarrow m_{n-1} < \frac{2Q+1-\sqrt{8Qc+1}}{2} \cup m_{n-1} > \frac{2Q+1+\sqrt{8Qc+1}}{2},
\end{aligned}
$$

but, as $\frac{2Q+1+\sqrt{8Qc+1}}{2} > Q$, this simply reduces to the first inequality, i.e., to

$$
m_{n-1} < \frac{2Q+1-\sqrt{8Qc+1}}{2}, \tag{8}
$$

that then provides the optimal threshold in (7).

We may observe that it is never optimal to check the last alternative for values of the unitary cost for investigation such that

$$
c \geq \frac{Q-1}{2}, \tag{9}
$$

as in this case, the right term of the inequality in (8) is always smaller than or equal to 1, and then the inequality is never verified.

Step t: As before, we suppose that the consumer has already checked the first t alternatives. Let M_t be the random variable maximum of the first t observations. The agent knows its realization m_t, as she knows the realization q_i of each random variable Q_i with $i \in \{1, \ldots, t\}$, results of all the past observations. Then, the proof follows exactly as for $n-1$, providing the same threshold which is then independent of the step t.

□

We observe that the optimal stopping rule strongly depends on the cost factor c. In particular, when the cost is really small, it is convenient for the agent to check for all the available alternatives. In fact, in such a case, the threshold turns out to be smaller than the minimum feasible value of m_t, and then the inequality is always verified. On the other side, while the cost for investigation is increasing, it is convenient for the agent to speed up in choosing an option. In fact, in such a case the threshold is on the right of the minimum feasible m_t. In our analysis, we will investigate the model for two unit costs: a low one, c_l, and a high one c_h, representing two different kinds of consumers. The first one is more exigent, and she dwells on reading deeply the reviews, whereas the second one gets a look and goes further. Assuming a uniform distribution of quality products is a good approximation when the consumer ignores the strategy that the platform has adopted for showing the goods. This assumption provides us with a threshold which is optimal in this specific context, but which may not be optimal when the consumer has more information about the possible distribution of the quality. What happens in reality, in fact, is that the platform defines an ordering method, in particular a decreasing rating based ordering, and a decreasing price based ordering. In these settings, the quality of the goods are then not uniformly distributed anymore, and the consumer could infer this information by the knowledge of the proposed ordering. Indeed, getting the distribution of the quality could be way too complicated for the consumer, who must devote time and attention for decoding all the information signals useful for a quality evaluation.

5.3 Optimal Search

In this section, we illustrate the consumer searching behaviour we adopt in our model. We suppose that the consumer enters the products page of an online retailer, and she is going to explore some items. In our setting, the consumer does not know exactly the product to search for, but she just has an idea about some product features, and she is submitted to a budget constraint $\bar{p}$. As final goal, she aims at finding a good quality product in a short time. Generally speaking, a user whose goal is refining the products' exploration, can make some actions: activating search filters (checking boxes of product attributes), consulting the retailer's recommendation system or querying the local search engine. Since, in our setting, the user does not known the product, the simplest way to convey his preferences, is switching *on* the attribute product filters for selecting the favourite sort. In our assumptions, compliant to the vast majority of the platforms, she has three by default ways of sorting the products: by price, by rating and random. We assume that the consumer learns prices and rating at a fixed cost for each item retrieved (as said before, we suppose two alternative unit costs: low one c_l and high one c_h). In this way, we can formulate a linear growth of global cost for searching.

Given a way of sorting the products, a consumer may terminate the searching as soon as she reaches an option whose price is below her budget constraint. Conversely, a consumer adopting a stopping rule is submitted to a more sophisticated decision mechanism. In particular, a consumer adopting our SR, terminates the searching as soon as both the price and the quality constraints are met (see Theorem 1). Our goal is to demonstrate that a consumer who searches a product and adopts our stopping rule, increments her expected utility with respect to what she could get without it. In fact, by means of the data contained in our dataset, we show that our mechanism is successful at increasing the average quality of product found.

It is worth noting that our results concern the product click rate[44], rather than the effective product purchasing action, on which we do not have data. However, as some literature points out (see Donnelly[26]), clicking and purchasing are two actions which are strongly correlated, to such an extent that the click rate can be adopted for both consumer surplus estimation and platform revenue ([26]).

[44]As Koulayev[32] underlines: "it is understood that the characteristic of good that affects how consumers search would also change how consumers click.[...] these changes are concerned with clicks on organic search results, whose quality is an important determinant of consumer satisfaction with the platform. Thus, both consumers and platforms owners are interested in maximizing the click rate on the platform".

The identification problem Let us show our optimal sequential search, introducing an *identification problem.* As Koulayev points out [32]) "the full identification of a search model requires that we are able to uniquely recover the joint distribution of preferences, search costs, and beliefs". About consumer tastes, we identify her wishes through the search mode (rating/price). Mimicking Koulayev "we assume that consumer beliefs can be reasonably approximated by the empirical distribution of product prices and qualities". In our assumptions, the consumer tries different search mode combination, compliant to those admitted by the platform. The goal is to maximize her utility, retrieving the best quality product in a minimum time, among all the attempts that she has tried. For this purpose, we begin identifying the search model provided by the SR, to which of the alternative model without the SR.

5.3.1 *Search with SR*

Recall that in defining our stopping rule, for an ordering $\{q_1, \ldots, q_n\}$ the aim was to find the optimal stopping time t (i.e., the threshold on the quality) which maximizes the utility function in equation (6). In the previous Section, we have proposed a solution to this problem assuming a default static sorting (i.e., when the ordering $q_1, \ldots, q_n$ was given randomly following a uniform distribution sorting). Now, we generalize this setting toward dynamic search specifications (i.e., when the ordering is chosen according to some specific criteria) and we apply the same stopping rule and check for its effects on this specific ordering of the qualities.

In particular, we assume that a consumer entering the platform can choose a search mode $s \in S = \{by\ Rating, by\ Price, Random\}$. For each specification of the variable s, the platform provides an ordering of the products, according to a function Sort:$S \to Q \times \cdots \times Q$. This function can be understood as a map between s toward an ordered tuple $\{q_1, \ldots, q_n\}$.

We assume a discrete population of agents $\{1, \ldots, K\}$, whose reservation prices $\bar{p}_k$ for each $k \in \{1, \ldots, K\}$ are uniformly distributed in $P = [p_{min}, p_{max}]$. This range represents the market price segment, within which it is carried out the search (eg.: see Section 3.3: Price segments). Similarly, each agent k is provided by one of two possible unit costs of searching $c_k \in C = \{c_l, c_h\}$.

Then, the utility function of a consumer k when she is selecting a product implementing our stopping rule is given by a function $U_k^{SR}(\bar{p}_k, c_k, s) : P \times C \times S \to \mathbb{R}$ that, given a tuple $(\bar{p}_k, c_k, s)$, assigns the corresponding value of the utility in equation (6) when the price threshold is given by $\bar{p}_k$, the cost of searching by c_k, the ordering of the products quality by Sort(s) and when the consumer stops searching when the price threshold is verified, and when the quality threshold in Theorem 1 is verified as well[45].

[45] If we try a more detailed comparison with Koulayev[32], despite the utility function of his work takes into accounts more arguments, its form can be reduced to: $u_{i_j} = \mu(p_j, not\_p_j) + \epsilon_{i_j}$ where p_j is the price of product j, not_p_j is the not-price characteristic of product j, and ϵ_{i_j} is the alternative shock interpreted as uncertainty of consumer i about the value of search (e.g., future search cost)[32]. Our function can be splitted in price and not-price values too

Fixing a reservation price $\bar{p_k}$, if we attempt all the specifications c_k and s and we maximize on these two variables, we obtain a distribution of utility values given by

$$u^*(p_k) = \max_{(c_k,s)\in C \times S} U_k^{SR}(\bar{p}_k, c_k, s)^{46}$$

.

In our model, we suppose each agent k will click on the product g_k corresponding uniquely to the utility $u^*(\bar{p_k})$ that make her reach the best quality result of all the specifications of c_k and s. We observe that $\forall\ \bar{p_k} \in P$ we obtain a value $u^*(\bar{p_k})$ that we will fit to a continuous function compliant to this discretization.

5.3.2 *Search without SR*

The description of this case is simpler. The utility function $U^{NOSR}(\bar{p_k})$ when the consumer is not implementing the SR, is given by the utility she gets when she purchase the first product which respect the price constraint $U^{NOSR}(\bar{p_k}) = \text{first}\{q_1, \ldots, q_t | p_t < \bar{p_k}\} - c_t$. The corresponding ${u^*}^{NOSR}(\bar{p_k})$ is calculated in the same way of SR case.

6 Estimation

The value $u^*(\bar{p_k})$, given a reservation price p_k is achieved maximizing a set of values, each of which is in turn the result of another maximum provided by the SR formula. Informally, we have achieved a set of maximal values, each of them is the best value retrieved among each search specification. The estimation of this class of problems is approached by literature through the Generalized Extreme Value distribution of Type-I (GEV-I), also called Gumbel distribution[47]. Practically the Gumbel is a distribution of a set of distributions. This type of distribution is adopted as closed-form expression for estimation utility function in optimal search problems by authors already mentioned here: Koulaev[32], De los Santos[15], Honka[23][48].

(the quality, which is none other than an aggregation of not-price attributes). About the term ϵ_{i_j} that estimates the uncertainty, our model does not provide this, because we consider that the SR applied to the quality absorbs endogenously any consumer incertitude.

[46] The existence of a unique value is not assured. However, the probability to find a higher number of maxima, when implementing it on our dataset, is negligible.

[47] The Gumbel distribution is used to represent the maximum or minimum of a number of samples of size n, of various distributions. Let us make an example: supposing we are interested in the biggest wave for a century, if there was a list of maximum waves for the past ten years. It is useful in predicting the probability that an extreme seaquake hits a bridge or a port. This distribution is studied in an area of statistics that is known as extreme value theory.

[48] In the wake of Koulaev[32] and Honka[23], we simulate the consumers' Utility through a Gumbel CDF distribution, simplifying their framework. Koulaev uses a combination of EV-I distributions, because he is interested in the utilities chain estimation of all the search attempts, until the stopping of decision mechanism. Honka adopts EV distribution of prices,

For each consumer k, provided by a reservation price $\bar{p_k}$, we can observe the best quality products $\hat{G} = \{\hat{g}_k\}$, each one retrieved by a search specification $\mathrm{Sort}_{(c,s)}\ (\bar{p_k}, \cdot, \cdot)$. We call this product the *clicked-price product*, because it is the best quality product reached by a reservation price. In order to extend the discrete list of consumers to a continuous distribution, we estimate the likelihood that a consumer k can reaches g_k, for all $p \in [p_{min}, p_{max}]$, then we fit $\hat{G}$ to a Gumbel cumulative distribution function $G(p)$. The function represents the maximum expected utility value that a consumer equipped with a reservation price p can click on g. That is:

$$u^*(p) = \mathbf{E}\big[\max_{g \in G}\{u_{(p,g)}\}\big] \tag{10}$$

The equation in (10) defines the consumer utility continuous function $u^*(p)$.

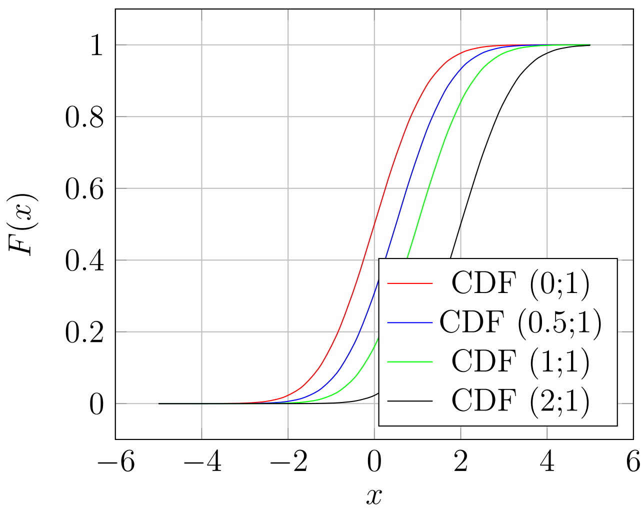
Gumbel distribution A random variable x has a Gumbel distribution if its probability density function PDF is defined by (11) and its cumulative distribution CDF defined by (12)

$$f(x|\mu, \sigma) = \frac{1}{\sigma} \exp\left\{\left(\frac{x-\mu}{\sigma}\right) - e^{\left(\frac{x-\mu}{\sigma}\right)}\right\} \tag{11}$$

$$F(x|\mu, \sigma) = 1 - \exp\left\{-e^{\left(\frac{x-\mu}{\sigma}\right)}\right\} \tag{12}$$

- This is a location-scale family of distributions, where μ is the location and $\sigma > 0$ the scale.

Graphical representation of Gumbel CDF, setting different values of (μ, σ), in red $\mu = 0, \sigma = 1$.



showing that the maximum utility function is in turn a EV distribution. In these models, the utility is a linear combination of price and not-price characteristics. Both the authors use a multivariate scaled logistic CDF to estimate the parameters.

Let $p_{g_1}, \ldots, p_{g_n}$ the prices of product instances of our sample G (as a particular case of $x_1, \ldots, x_n$ random sample), the "Log Maximum Likelihood" function of CDF is given by the equation below[49].

$$\mathcal{L}(\mu, \sigma) = -\sum_{g_j \in G} \left(\frac{p_{g_j} - \mu}{\sigma} \right) - \sum_{g_j \in G} \exp \left\{ - \left(\frac{p_{g_j} - \mu}{\sigma} \right) \right\} \tag{13}$$

The parameters estimation is obtained by:

$$\hat{\mu} = -\hat{\sigma} \log \left[\frac{1}{n} \sum_{g_j \in G} \exp \left(-\frac{p_{g_j}}{\hat{\sigma}} \right) \right]$$

$$\hat{\sigma} = \bar{g} - \frac{\sum_{g_j \in G} \exp \left(-\frac{p_{g_j}}{\hat{\sigma}} \right)}{\sum_{g_j \in G} \exp \left(-\frac{p_{g_j}}{\hat{\sigma}} \right)}$$

In our case, (μ, σ) are respectively the location and the common scale of the observed prices.

6.1 Results

We explicitly apply the search process, as it is defined in Section 5.3, to various set of data. The process is drawn by a Sort function $\text{Sort}_{(c,s)}\ (\bar{p_k}, \cdot, \cdot)$ as (c, s) change: $c \in C = \{c_l = 0.07, c_h = 0.15\}$, $s \in S = \{\textit{by Rating},$ *by Price*, *Random*$\}$ and reservation price $\bar{p}_k \in P = \{\bar{p}_1, .., \bar{p}_K\}$ uniformly distributed in $[p_{min}, p_{max}]$. In our simulation the enumeration of all sets is: $c = 2$, $\#s = 3$, $k \in \{1, .., 100\}$. The two borders $[p_{min}, p_{max}]$ represent the market price segments {Low Price, Medium-Low Price, Medium-High Price, High Price} described in (Section 3.3). For each segment the magnitude of all combination is #600, for all segments is #2400.

6.1.1 *Benefit and Search Cost*

Our scope is to estimate which is the optimal quality product, its price p_g^* and the search cost $\#steps$ retrieved by the best search matching. We investigate around four product price segments of Camera category drafted before (eg.: see Section 3.3: Price segments), considering two search specifications: $\text{Sort}_{(\{c_l, c_h\}, s)}$ with SR and without. In each of which, the consumer will pick the

[49]All the code used for this paper is available from (`"https://github.com/maroccasting/economics-experiments"`).
For the estimation, we use a package python scipy.stats focused on distribution fitting. Among various MLE fitting algorithms, we adopt the implementation of method of the moments. This implementation gives a goodness of fit.

product $g \in G$ provided by the largest benefit, as it is defined in equation(10), giving the reserved price $\bar{p_k}$. We remember that this product satisfies the max quality reached in the range $[p_{min}, \bar{p_k}]$. The Gumbel CDF in equation (12) is the distribution that we are going to fit. MLE of Gumbel CDF distribution maximizes the probability to be compliant to the values observed, estimating the best parameters location μ and scale σ, as in equation (13). In the first 2 columns of table 10 we show the location estimation μ and the st. error for each specification with SR and without. Comparing the location parameter of the fitted CDF[50] SR driven $p^*_{g^{SR}}$ with the alternative without SR $p^*_{g^{NOSR}}$, in the first case the price is lower than the second one. This implies that the consumer, through the SR mechanism, can reach a more quality cheaper product with more probability.
We show also the quality and the search cost of such a product. The difference between quality product SR driven and the alternative without, is always $> 7\%$ in each price segment, reached at a lower average price. We observe the best SR contribution, in terms of quality and search cost, is reached for Medium-High price segment. Whereas for cheapest products the absolute benefit is lower and the cost highest. The higher price difference ($p^*_{g^{SR}}$ - $\bar{p}$) and the higher quality difference ($q^*_{g^{SR}}$ - $\bar{q}$) with respect to the median, give an idea of the contribution of the tool with respect to a simple sorting by price. For a panorama of detailed attempts see the table 15 at page 84 and the table 16 at page 85.

Table 10: Price, Search cost and Quality of optimum Clicked-product under different specification

Price segment	**clicked Price** (\$)		**Search cost**		**Quality**	
	Sort^{SR}	Sort$^{NO\ SR}$	Sort^{SR}	Sort$^{NO\ SR}$	Sort^{SR}	Sort$^{NO\ SR}$
Low	15.7 (0.04)	14.1 (0.03)	*6*	*4*	*0.88*	*0.75*
Medium-Low	51.2(0.09)	54.3 (0.05)	5	**2**	0.9	0.85
Medium-High	155.2 (0.08)	164.5 (0.04)	**3**	**2**	**0.93**	**0.85**
High	432.4 (0.09)	411.8 (0.11)	4	**2**	0.9	0.83
# sample	600	600	600	600	600	600
Log likelihood	-38.75	-37.52				
p-value (mean)	0.07	0.08				

Notes: The clicked price is an optimal estimated value. The search cost and quality are referred to the actually close product. Search cost is (# steps), Quality is normalize to 1.

[50] The fitted CDF is the price $p^*_{g^{SR}}$ for which the consumer optimizes its benefit. This value is an estimated value, involving the ideal product close to the actual best quality one. From a statistical p.o.v. the location is the argmax of PDF density function.

Table 11: Median of Price, Rating and Quality values for each product segment

Price segment	Price ($) mean (sd)	Rating SR	Rating NoSR	Quality SR	Quality NoSR
Low	16 (7.4)	4.4	4.3	*0.82*	*0.56*
Medium-Low	54 (21)	*4.3*	4	0.83	0.75
Medium-High	164 (35)	4.4	*4.2*	**0.92**	**0.78**
High	598 (220)	**4.6**	**4.6**	0.85	0.75

Notes: Search cost (# steps), Rating, Quality column with and without SR. Mean values [High unit cost c_h=0.15]

6.1.2 *Stocastic Dominance*

The simulation of the model has empirically shown that, given an i.i.d distribution of agents, facing two specifications of search modes, Sort^{SR} and $\text{Sort}^{\text{NOSR}}$, the benefit of the best quality product reached by SR is higher than the alternative $NOSR$. However, for each price segment, the result is obtained only for the argmax of the density function (the first derivative of our utility distribution function), and does not concern the full range of prices $[p_{min}, p_{max}]$. To extend the point-wise result to the domain level, we are going to estimate the first-order dominance of SR specification comparing the alternative.
Although in economics the notion of stochastic dominance is generally referred to maximizing expected utilities between two or more lotteries, we can try to apply it to the distribution utility function proposed in the previous Section. the Stochastic dominance is a partial order between random variables, useful to make a decision over some preferences. In light of this, giving two generic distribution functions $F(x)$ and $G(x)$, a decision maker that observes a First-order stochastic dominance, within a range of $x \in [x_{min}, x_{max}]$ can make a decision formally based on the best benefit received.

First-order dominance definition. *For any distribution function $F(x)$ and $G(x)$, $F(x)$ first-order stochastically dominates $G(x)$ if $F(x) \geq G(x)$ $\forall x$.*

We apply the *Kolmogorov-Smirnov* test[51], to verify for which price segment

[51] The python scipy.stats implementation specifies the null hypothesis that the true distribution function of x is equal to, not less than or not greater than the distribution function of y (two-sample case). Like in R implementation.
(`https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ks_2samp.html#scipy.stats.ks_2samp`)
(`https://stat.ethz.ch/R-manual/R-devel/library/stats/html/ks.test.html`)

the SR distribution dominates the $NOSR$ distribution. Supposing a significance level $\alpha = 0.01$. We find for low-level price segment a p-value = 0.002, for medium-high price segment p-value = 0.001 and for high price segment a p-value = 0.005. Since $p < \alpha$ in this three case we can reject the null hypothesis. Whereas in low-high price segment p-value = 0.023 then we cannot reject the null hypothesis. In fact for some product belonging to this segment, the utility value is higher without SR, as reported in same case of the table 15 at page 84 and the table 16 at page 85.

6.1.3 *Price elasticity of demand*

The literature with which we compare our model, simulates the *price elasticity of demand*, by measuring the variation in volume of retrieved products (by a search mechanism specification) in response to a variation in price[52]. According to Kim et al.[31], in our model, the own-price elasticity of demand (ϵ^{search}) is seen as a percentage change in expected click rate of retrieved products following one percent increase in price. Practically, we predict the volume of retrieved products through different consumer search settings, as the marginal price changes. Our scope is to evaluate, essentially in which way, and for which segment of prices, the SR affects the demand's elasticity.

$$\epsilon^{search} = \frac{dQ_{ty}/Q_{ty}}{dp/p} \tag{14}$$

Where $p^{+1\%}$ represents an increment of 1 unit in price and $Q^{p+1\%}$ the comparative quantity affected by the price increment, $dQ_{ty} = Q_{ty} - Q_{ty}^{p+1\%}$ and $dp = p - p^{+1\%}$. Given a product i the definition ϵ_i^{search} is applied to this product. We observe that our measure concerns the **own-price** elasticity of demand of single products. Whereas the **cross-price** elasticity of demand shows how quantity demand measure the effects of related goods. To evaluates these effects, it needs a basket containing more closely substitutable products, difficult to collect in our dataset.

Resuming the conditions of search setting (see Section 5.3: Optimal Search), we explain exactly how the quantity of products is calculated. Giving a consumer k provided with a reservation price $\bar{p_k}$ and two search specifications: $\text{Sort}_{(\{c_l,c_h\},s)}$ with SR and without. In each of which, the consumer will pick the product $g_i \in G$ provided by the largest benefit. we define the cardinality of the product $\#\{g_i\} = \sum_{k=1}^{K} \#g_{i_k}$ as the amount of occurrences are retrieved as k changes. According to the *price elasticity of demand* definition, it easy

[52]In De los Santos[15], the formula is more complicate, but essentially the author calculates the average change in buying probability of a marginal increasing of price (taking in account either the product price and the average of the expected price distributions in the marketplace).

to see that the quantity $Q_{ty}(i)$ in (14) is the cardinality $\#\{g_i\}$[53]. To take an example, ϵ_i= -0.5 means that if we increase the price of product i by 1%, its search quantity will drop by 0.50%. We apply the same experimental setting used in Section 6.1: Results[54].

Table 12: Price elasticity of demand (low price)

Product Name	**Brand**	**Price** ($)	**Rating**	**Quality**	ϵ_i^{SR}	$\epsilon_i^{NO\,SR}$
Wasabi Power Battery (2-Pack)	Wasabi	21	4.5	0.85	-0.9	-1.3
Mediabridge Ultra Series RCA	Mediabridge	11	4.4	0.85	-0.8	-1
Micra Digital CAT5 Cable	Belkin	11	4.4	0.22	-2.1	-3.5
Anker® Golden AC Adapter	Anker	18	4.0	0.45	-2.0	-2.3

Table 13: Price elasticity of demand (high price)

Product Name	**Brand**	**Price** ($)	**Rating**	**Quality**	ϵ^{SR}	$\epsilon^{NO\,SR}$
Pentax K-5 16.3 MP	Pentax	370	4.8	1.0	-0.2	-0.4
Canon G12 10 MP Digital Camera	Canon	598	4.4	0.82	-0.5	-0.5
Nikon D3100 DSLR Camera	Nikon	440	4.6	0.6	-0.1	-0.4
Canon SX30IS 14.1MP Digital Camera	Canon	450	4.0	0.77	-0.3	-0.3

Generally speaking, the results on price elasticity of demand point out to the product quality and brand incidence in click rate. The SR affects only the results for low price products demand (see table 12 at page 75). Taken into exam two products, *Mediabridge Ultra Series RCA* and *Micra Digital CAT5*, the first one is a cable that has remained popular for 8 years until 2014, whereas the second one has disappeared in few time, due to obsolete technology, despite having the same rating. The elasticity of demand index, in the first case (with and without SR), is relatively inelastic. Consumers that have purchased the product still do if the price increases. Whereas, *Micra Digital CAT5*, with low quality and higher negative elasticity has been exposed to the pitfalls of the market. In the table 13, are illustrated high price products. Comparing with the previous group, the quality in average is higher, and the elasticity in average lower. In this segment, the brands are known and the brand loyalty assures small changes in selling, when the price changes. In table 14 at page 76 we have compared the products by price segments, taking into account the average elasticity in each segment. At Medium-High price level, as we have already seen before, quality is higher and search cost lower (because the

[53]The cardinality of $\{g_k\}$ is affected by the change of reserved prices, because some queries do not matching, whereas other ones retrieve the same product.

[54]The values of price data are referred to the 2014, the last year of running the scraping process of our dataset. All the values are referred to this period and not updated now.

probability to reach early the maximum quality product increases). This flat effect reduces in this task the difference between SR and NO SR. When the range of prices confines the luxury segment of the market category, people are disposed to buy the loyal products, even if the price increases. The demand is relatively inelastic (-1 $< \epsilon <$ 0) with an average value of -0.5 with SR and -0.35 without. At the contrary, in accessories segment, the stickiness to brand awareness fails, and the consumer makes more attention to price increasing, except in the cases of high quality products.
We confirm the results obtained by Kim et al.[31], a similar work focused to a very close market category (camcorder). For example, the Sony camcorder with DVD media format, 40×optical zoom, selling at $360 has a $\epsilon = -2.12$, where the cheapest product ($32) has $\epsilon = -0.58$.

Table 14: Price segments comparison

Price segment ($)	**Elasticity** ϵ	
	SR	NO SR
Low price $<$ 30	-2.0	-2.4
30 $\leq$ Medium-Low Price $<$ 120	-1.4	-1.5
120 $\leq$ Medium-High Price $<$ 250	-0.7	-0.7
High Price $\geq$ 250	**-0.5**	**-0.35**

Notes: in each line the average elasticity among the products of the same segment.

7 Discussion and Conclusion

Discussion and comparison with other models

In this article, we have described a model of consumer search for products, exploiting a decision-making mechanism based on a Stopping Rule (SR). The SR permits choosing the highest quality product in a sequential search between n alternatives, with respect to the minimization of search costs. Unlike many sequential search models, our implementation does not merely illustrate a consumer behaviour, but can make a decision, replacing the incertitude of the consumer. In other words, it can be used as a platform design tool for helping the consumer in their choices. There are two main differences between our model and the classical sequential search models, proposed by McCall[36]. First of all, our aim is to maximize quality rather than minimizing the target price, which is the aim of classical models where the product is known. Secondly, the vast majority of the models proceeds by attempts, evaluating in each step the difference between next expected utility and reserved ones already estimated, whereas our design terminates after a one-stage decision.

Quoting Kim et al. "the theory of optimal sequential search states that consumers only continue to search if the marginal benefits of doing so outweigh the marginal costs"[31]. Specifically, in the classical models, given a distribution of prices $F(p)$ and a reservation price p_r, a consumer chooses a specification to optimize the number of steps (search) to find a price close to p_r. Following this process, the consumer can formulate expectations about new price quote (the benefit) and evaluate uncertainty regarding the search (e.g.: future search marginal cost). Such a goal is achieved by gathering additional price information to determine the expected benefit of search. Practically, by means of a first attempt, the consumer obtains a minimal amount of knowledge about price distribution, and actualizes the belief of alternative events. The recent approaches are different because they take into account not only known products, but also unknown ones. This behavior is more compliant to consumer attitude in searching on platforms, because the daily growth of products makes for detailed knowledge about unknown new items difficult to discern. Searching new products shifts consumer attention to maximizing utility instead of minimizing prices. The studies interested in this new fuzzy consumer knowledge, apply a structural change for sequential search. The step-wise inference engine must be oriented to alternatives to the exact product name given that this is lost in this kind of mechanism. The most attractive are the not-price product attributes, such as those used by Kim et al.[31] and Koulayev[32][55]. One of the best models that has embarked upon this new style, is developed by Koulayev[32]), in which the author estimates a joint distribution between product characteristics and consumer taste, over more than twenty attributes. Concerning our work, the evaluation of the best consumer utility is based on a SR mechanism, which stops at a good enough choice (under a budget constraint). The stopping decision is made when the best quality item, with respect to the search cost, is reached. The quality index is able to digest not-price attributes, after evaluating them through an exogenous statistical learning algorithm. In the light of this solution, the stopping mechanism needs only a list of past value observations of quality signals, and a condition rule about the expected new alternatives. The target argument is a value, not a complex function. This value concentrates the quintessence of information quality. In the First Chapter, we have described its formulation process, starting with a function that includes many variables, that has been reduced to a simple value of product quality. The choice turned out to be good in many cases, as we have shown in our results. The critical conditions are represented by the heterogeneous distribution of price and information, that happens when we enlarge the searching domain. In the worst case, through One-shot attempt, some higher quality products are ranked after the stopping step. Under these

[55] To have an idea, two classes of attributes are taken into account: one is composed by product features, such as those made available by consumer reviews collections like *Consumer Reports Magazine*: weight, design, size, reliability, usability, and so forth. Another class is formed by check box attributes for filtering, generally placed on the left of platform product pages. Other than the rating, and other common features, depend on the specific product category (eg.: in the case of hotels: facilities, room service, panorama, breakfast, cleaning, comfort and so forth). *Consumer Reports Magazine* is an American nonprofit consumer organization dedicated to unbiased product testing. It provides the latest ratings and reviews plus rigorous reporting on issues with worldwide impacts. The vast majority of online platforms has followed this setting, exposing product information through UCG.

circumstances, alternative methods are to be preferred, specifically those that estimate the best utilities by trial and check. However, the heterogeneity can be reduced, by setting a partition of price segments like our experiment design. This condition, adopted by other authors such Kim et al.[31], does not limit idiosyncratic consumer taste, because searching for products over a large price range is improbable.

Conclusion

An important contribution of our research is investigating in price differentiation for searching. In particular, for different price segments: low priced products (accessories), medium priced (the target market of the camera category) and high priced ones (eg.: luxury market), with respect to product quality. We discover that, in the range of prices where the information is dense and more uniformly distributed among products, like medium-high price in camera electronic market, the rating of known brands remains a good proxy for the quality index, as predicted in most of the literature. This standard assumption is not true for accessories, for which the rating is not the best measure of quality. The greatest usefulness of the SR arises exactly in this segment. Finally, we have explored the quantitative implications of our proposed model by tracing the cost distribution within price segments. In medium-high level, the SR benefits in terms of reducing costs are proportional to advantages in terms of average product quality. The difference between the two settings, with SR and without, is 17% in best product quality. While, for the lowest price, the best benefit in quality difference between SR and No SR is higher 26%. The result has been achieved by paying a higher search cost. We have also estimated the price elasticity of demand with and without SR, by changing different parameters and search conditions. In particular, the SR offers visibility to the unbranded tools and components that still play a relevant role to support new design in searching tools. The estimates of utility parameters under different specifications highlights the search by rating with SR, as the best tool to capture optimal product quality. From the "Information Retrieval" point of view, this improvement can be seen as a re-ranking solution, potentially deliverable through an external add-on. We may conclude that, in many of the cases, the SR guarantees an higher utility, with respect to the traditional cases. This result is reflected by the shape of the utility curve, which dominates for all the reservation prices. The best result is achieved in the medium-high price group (100-240$), because it is the most competitive niche camera category, which shows the best average quality and proportionally best rating, and where the reviewers write the most interesting reports.
Definitively, we have crossed three different branches of study: *(i)* First of all, the Consumer Search on platform, which is the main research line of this paper, focused on a new model oriented toward reaching the optimal consumer utility *(ii)* Stopping Rule mechanism, typical of exploration-exploitation trade-off (see March[34]), in particular the *Secretary problem* (see Ferguson[19]) and *(iii)* Statistical learning algorithm concerning the evaluation of quality product through reviews content. The third, discovered and illustrated in the First

Chapter, turns out to be the ideal ingredient for a menu that mixes computer science research with economics of information.

Future research

Our work still has two limitations: the fact the sample is composed of only 100 elements and the unique market category of our sample. As future research we underline the need of generalizing our findings by applying our framework to other categories, such as book markets, and on a much larger sample.

Another interesting improvement would be to study the performance of the SR mechanism when assuming decreasing costs (because the consumer, after some steps, has already formed a rough idea about the reviews content, and we can assume it would be less costly for her to get more information in her search while time is passing).

Towards a perspective of market design evaluation, a future research concerns to describe a complete model of supply and demand. This model may be represented applying the tools of Game Theory, via a strategic setting in which n sellers (each one produces only one product) compete by choosing a strategy in the form of a pair (*price*, *quality*). In such a setting, when a platform decides to propose new products sorting (eg.: implementing our quality index), the sellers can react strategically and modify the price or the quality, in order not make their position worst, in the proposed sorting. In this condition, the consumers do not act strategically, and they have only the passive role of observing the search results and making the suggested choice. Assuming a dynamic setting, suppose a seller who strategically decreases the price or increases the quality in order to become more competitive. The other sellers, as competitors, can react by trying the best response to such a change. Such a game is interesting for evaluating price differentiation and total revenues of sellers (estimated by the click rate of first page position) on different equilibria.

References

[1] Baye, Michael R., Babur De los Santos, and Matthijs R. Wildenbeest. "The evolution of product search." JL Econ. & Pol'y 9 (2012): 201.

[2] Baye, Michael R., John Morgan, and Patrick Scholten. "Price dispersion in the small and in the large: Evidence from an internet price comparison site." The Journal of Industrial Economics 52.4 (2004): 463-496.

[3] Baye, Michael R., John Morgan, and Patrick Scholten. "Information, search, and price dispersion." Handbook on economics and information systems 1 (2006): 323-375.

[4] Bhargava, Hemant K., and Juan Feng. Does better information lead to lower prices? Price and Advertising Signaling under External Information about Product Quality. working paper, 2015.

[5] Bodoh-Creed, Aaron, Jörn Boehnke, and Brent Hickman. Using Machine Learning to predict price dispersion. Working Paper, 2018.

[6] Blake, Thomas, Chris Nosko, and Steven Tadelis. "Consumer heterogeneity and paid search effectiveness: A large scale field experiment." Econometrica 83.1 (2015): 155-174

[7] Brynjolfsson, Erik, and Michael D. Smith. "Frictionless commerce? A comparison of Internet and conventional retailers." Management science 46.4 (2000): 563-585.

[8] Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith. "From niches to riches: Anatomy of the long tail." Sloan Management Review 47.4 (2006): 67-71.

[9] Chevalier, Judith and Dina Mayzlin 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." Journal of Marketing Research 43(3): 345-354.

[10] Coad, Alex. "On the distribution of product price and quality." Journal of Evolutionary Economics 19.4 (2009): 589-604].

[11] Crosby, Philip (1979), "Quality is Free" New York and Scarborough, Ontario: Mentor Books

[12] Curry, David J. and Peter C. Riesz. (1988), "Prices and Price/Quality Relationships: A Longitudinal Analysis." Journal of Marketing 52, no. 1, pages: 36-51. doi:10.2307/1251684

[13] DellaVigna, Stefano, and Matthew Gentzkow. "Uniform Pricing in US Retail Chains." Available at SSRN 3367978 (2019).

[14] De los Santos, Babur, and Sergei Koulayev. "Optimizing click-through in online rankings for partially anonymous consumers." Marketing Science (próxima publicación) (2012).

[15] De los Santos, B., Hortacsu, A., and Wildebeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. The American Economic Review, 102(6) : $2955-2980$

[16] De Los Santos, Barbur, Ali Hortaçsu, and Matthijs R. Wildenbeest. "Search with learning for differentiated products: Evidence from e-commerce." Journal of Business & Economic Statistics 35.4 (2017): 626-641.

[17] Dinerstein, Michael, et al. "Consumer price search and platform design in internet commerce." American Economic Review 108.7 (2018): 1820-59.

[18] Ellison, Sara Fisher. "capt.12 Price search and obfuscation: an overview of the theory and empirics" Handbook on the Economics of Retailing and Distribution (2016): 287.

[19] Ferguson, Thomas S and others, "Who solved the secretary problem?", Statistical science 4(3) (1989): 282-289

[20] Fisher, Matthew, George E. Newman, and Ravi Dhar. "Seeing stars: How the binary bias distorts the interpretation of customer ratings." Journal of Consumer Research 45.3 (2018): 471-489.

[21] Hayes, Robert H. and Steven C. Wheelwright (1984) "Restoring Our Competitive Edge". New York: John Wiley & Sons, Inc.

[22] Hey, J.D., "Still searching", Journal of Economic Behavior & Organization 8 (1987): 137–144

[23] Honka, Elisabeth. 2010. "Quantifying Search and Switch- ing Costs in the U.S. Auto Insurance Industry." http://home.uchicago.edu/ ehonka/Paper EHonka 100310.pdf.

[24] Honka. 2014. Quantifying Search and Switching Costs in the US Auto Insurance Industry. The RAND Journal of Economics 45(4), 847-884.

[25] Morris, Stephen, and Philipp Strack. 2017. "The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs." Working paper, Princeton Univ.

[26] Donnelly, Robert, Ayush Kanodia, and Ilya Morozov. "A Unified Framework for Personalizing Product Rankings." Available at SSRN 3649342 (2020).

[27] Ellison, Glenn, and Sara Fisher Ellison. "Search, obfuscation, and price elasticities on the internet." Econometrica 77.2 (2009): 427-452.

[28] Ellison, G. and S. Fisher Ellison (2014), 'Match quality, search, and the Internet market for used books', working paper, Massachusetts Institute of Technology (MIT), Cambridge, MA.

[29] Einav, L., Farronato, C., Levin, J. (2015). "Peer-to-Peer Markets". National Bureau of Economic Research Working Paper Series, No. 21496. doi: 10.3386/w21496

[30] Karlis, Dimitris, and Evdokia Xekalaki. "Mixed poisson distributions." International Statistical Review 73.1 (2005): 35-58.

[31] Kim, Jun, Paulo Albuquerque, and Bart J. Bronnenberg. 2010. "Online Demand Under Limited Consumer Search." Marketing Science, 29: 1001-1023.

[32] Koulayev, S. (2014). Search for differentiated products: identification and estimation. RAND Journal of Economics, 45(3):553-575.

[33] Bertin Martens (2016) "An Economic Policy Perspective on Online Platforms". Institute for Prospective Technological Studies Digital Economy Working Paper 5 (2016).

[34] March, James G, "Exploration and exploitation in organizational learning", Organization science, 2(1) (1991): 71-87.

[35] Moraga-González, José L., Zsolt Sándor, and Matthijs R. Wildenbeest. "Consumer search and prices in the automobile market." (2015).

[36] McCall, John J. 1970. Economics of Information and Job Search." Quarterly Journal of Economics, 84: 113-126.

[37] Seale, D.A., Rapoport, A., "Optimal stopping behavior with relative ranks: The secretary problem with unknown population size", Journal of Behavioral Decision Making 13 (2000): 391–411.

[38] Spence, F. (2014), 'Consumer experience and the value of search in the online textbook market', working paper, University of Notre Dame, South Bend,

[39] Stigler, George. 1961. The economics of information. Journal of Political Economy. 69(3) 213-225.

[40] Zwick, R., Rapoport, A., Lo, A.K.C., Muthukrishnan, A., "Consumer sequential search: Not enough or too much?", Marketing Science 22 (2003): 503–519

A Parameters Comparison under different specification

In table 15 and 16 at pages 84 and 85 we described search cost (number of steps), rating and product quality, calculated in the three searching mode: (*by Price*, *by Rating* and *Random*), at the two unit cost of search in the two cases: SR and NoSR. All these valued are referred to the median reservation price of each segment. The results with SR show higher product quality with respect to the alternative without (in only one case among 15 the stopping rule performs worse).
Search *by Rating* represents the basic proof of measuring the performance of our stopping rule implementation, because put in evidence that the rating system is not always a good index for quality evaluation. In fact, for prices less than 100$ (in which the variance of rating is higher, and the mean lower) the rule confirms an increment in average quality maintaining the same level of rating, as shown in table 15 at page 84. In the case of Medium-high and High prices the rating is directly proportional to the quality. In Search *by Price* specification, without stopping rule the result achieved is the bottom product price. Adopting the Stopping Rule. it is a balance between the basic order *by Price*, and the more priced item providing a not inferior quality.
The tables 17 at page 86 and 18 at page 86, summarize the results of the big tables 15 and 16. We observe that, for low price and medium-low, the rating is around 4.3 both for SR and No SR, but for low price, the quality of SR is 0.30 higher, whereas for Medium-low one, the quality of SR is 0.11 higher. That put in evidence the weak reliability of rating in the accessories segment. Changing the unit cost, from c_l=0.07 to c_h=0.15, the results with SR change a bit. The search cost decreases (e.g.: for Low Price the average number of steps turns from 6 to 4.2), but also the quality. **Note** As the rating of product, we assume the average rating of all the reviews. It is important to observe that this is not exactly what a platform usually does, when they show a rating value, but it may differ of some points percentage from the other one. For details see Section Related literature of Chapter 1.

Table 15: product-clicked values of parameters in details. Low cost unit

Segment level Price	Order	# items shown	reservation price ($)	Search cost (# steps)	Price, Rating, Quality	Stopping rule
Low price	By Price	20 of 25	20	8	11$, 4.4, 0.85	Y
				1	11$, 4.4, 0.85	N
	By Rating	20 of 25	20	4	11$, 4.4,0.85	Y
				1	11$, 4.5,0.22	N
	Random	20 of 25	20	6	11$, 4.4,0.85	Y
				1	10$, 4.0,0.57	N
Medium-Low price	By Price	20 of 25	70	5	35$, 4.3,0.95	Y
				1	30$, 4.6,0.73	N
	By Rating	20 of 25	70	4	35$, 4.3,0.96	Y
				1	30$, 4.6,0.73	N
	Random	20 of 25	70	4	40$, 4.1,0.85	Y
				2	67$, 3.5,0.65	N
Medium-High price	By Price	20 of 25	200	3	120$, 4.0,0.99	Y
				1	110$, 4.4,0.70	N
	By Rating	20 of 25	200	1	157$, 4.6,1.0	Y
				1	157$, 4.6,1.0	N
	Random	20 of 25	200	6	160$, 4.3,1.0	Y
				3	179$, 3.4,0.64	N
High Price	By Price	20 of 25	700	9	450$, 4.0,0.77	Y
				1	270$, 4.0,0.52	N
	By Rating	20 of 25	700	4	370$, 4.8,1.0	Y
				2	305$, 4.9,0.43	N
	Random	20 of 25	700	3	540$, 4.8,0.89	Y
				1	454$, 4.9,0.50	N
All Prices	By Price	20 of 25	250	3	9$, 4.0,0.76	Y
				1	5$, 3.4,0.34	N
	By Rating	20 of 25	250	4	11$, 4.4,0.87	Y
				2	40$, 4.6,0.65	N
	Random	20 of 25	250	2	25$, 4.2,0.80	Y
				1	10$, 4.0,0.59	N

Notes: # Steps, Rating, Quality column with and without SR. Mean values [High unit cost c_h=0.15]

Table 16: product-clicked values of parameters in details. High cost unit

Segment level Price	Order	# items shown	reservation price ($)	Search cost (# steps)	Price, Rating, Quality	Stopping rule
Low Prices	By Price	20 of 25	20	4	9$, 4.0, 0.75	Y
				1	11$, 4.4, 0.85	N
	By Rating	20 of 25	20	4	11$, 4.4,0.85	Y
				1	11$, 4.5,0.22	N
	Random	20 of 25	20	6	11$, 4.4,0.85	Y
				1	10$, 4.0,0.57	N
Medium-Low price	By Price	20 of 25	70	1	30$, 4.6,0.74	Y
				1	30$, 4.6,0.73	N
	By Rating	20 of 25	70	4	35$, 4.3,0.96	Y
				1	30$, 4.6,0.73	N
	Random	20 of 25	70	4	40$, 4.1,0.85	Y
				2	67$, 3.5,0.65	N
Medium-High price	By Price	20 of 25	200	1	110$, 4.4,0.70	Y
				1	110$, 4.4,0.70	N
	By Rating	20 of 25	200	1	157$, 4.6,1.0	Y
				1	157$, 4.6,1.0	N
	Random	20 of 25	200	6	160$, 4.3,1.0	Y
				3	179$, 3.4,0.64	N
High price	By Price	20 of 25	700	5	314$, 4.5,0.70	Y
				1	270$, 4.0,0.52	N
	By Rating	20 of 25	700	4	370$, 4.8,1.0	Y
				2	305$, 4.9,0.43	N
	Random	20 of 25	700	3	540$, 4.8,0.89	Y
				1	454$, 4.9,0.50	N
All Prices	By Price	20 of 25	250	3	9$, 4.0,0.76	Y
				1	5$, 3.4,0.34	N
	By Rating	20 of 25	250	4	11$, 4.4,0.87	Y
				2	40$, 4.6,0.65	N
	Random	20 of 25	250	2	25$, 4.2,0.80	Y
				1	10$, 4.0,0.59	N

Notes: Result details: # Steps, Rating, Quality column with and without SR. Mean values [High unit cost c_h=0.07]

Table 17: product-clicked values of parameters. Low cost searching

Price segment	**Price** (\$)	**Steps**		**Rating**		**Quality**	
	mean (sd)	SR	NoSR	SR	NoSR	SR	NoSR
Low	16 (7.4)	*6*	**1**	4.4	4.3	*0.85*	*0.56*
Medium-Low	54 (21)	4.2	1.3	*4.2*	4	0.88	0.69
Medium-High	164 (35)	**3.3**	*1.7*	4.3	*4.2*	**0.99**	**0.78**
High	598 (220)	5.3	1.3	**4.6**	**4.6**	0.88	0.59

Notes: # Steps, Rating, Quality column with and without SR. Mean values [Low unit cost c_h=0.07]

Table 18: product-clicked values of parameters. High cost searching

Price segment	**Price (\$)**	**Steps**		**Rating**		**Quality**	
	mean (sd)	SR	NoSR	SR	NoSR	SR	NoSR
Low	16 (7.4)	4.2	**1**	4.4	4.3	*0.82*	*0.56*
Medium-Low	54 (21)	4.0	1.3	*4.3*	4	0.83	0.69
Medium-High	164 (35)	**3.1**	*1.7*	4.4	*4.2*	**0.92**	**0.78**
High	598 (220)	*5.0*	1.3	**4.6**	**4.6**	0.85	0.59

Notes: # Steps, Rating, Quality column with and without SR. Mean values [High unit cost c_h=0.15]

Chapter 3

Price Matching and Platform Pricing

Abstract

In this study we investigate the effects of Price Matching Guarantees (PMG) commercial policies on U.S. online consumer electronics daily prices. By applying a Diff-in-Diff identification strategy we find evidence in favor of price reductions occurring after the PMG policy is repealed.
We further investigate if such effect is heterogeneous according to products characteristics, by exploiting User Generated Contents (products popularity and quality) and online search visibility measures (Google Search Rank). Estimates suggest that for high quality (visibility) products PMG policies harms competition by keeping prices high, while for low quality (visibility) products, prices decrease during the policy validity period.

1 Introduction

Online sales platforms have recently gained increasing importance in both retail and wholesale markets.[56] Such markets are characterized by the supply of personalized services, more convenient delivery schedules and the ability to reach a very high number of consumers. In addition, platforms claim to warrant lower prices with respect to traditional stores through the provision of offers, promotions, down prices and other price discounting policies. Among these options, online sales platforms often implement Price Matching Guarantees policies (PMG), that is the promise to reimburse price differences when competitors offer a lower prices.[57]

PMG policies are surely appealing for customers and can increase consumer confidence and brand fidelity. However, the announcement to tie prices to those of competitors can have anti-competitive effects and sustain high prices, thus harming consumers welfare.

[56] The term “online platform” identifies a range of digital services that facilitates interactions between two or more distinct but interdependent sets of users (whether firms or individuals) who interact through the service via the Internet [31]. Online sales platforms can operate as online retailers, as a marketplace for third-party sellers or they can offer both services.

[57] For example, NewEgg PMG policy states that "if you purchase an item from Newegg.com which is carrying the Price Match Guarantee badge at the time of purchase, then find the exact same item at a lower price by Newegg or a major retailer, just let us know, and we'll send you a Newegg Customer Care Card to cover the difference". See `https://kb.newegg.com/knowledge-base/price-match-guarantee/`.

Most of theoretical literature agrees on the fact that PMG reduce firms incentive to compete on prices and lower the motivation for consumers to search better sale conditions [20; 34]. However, in some models, PMG are considered as tools for price discriminating or as real discounting policies [5; 32]. Therefore, empirical analyses become particularly relevant in order to understand under what conditions such pricing policies reduce consumer welfare. Indeed, the applied literature analyzing this issue is scant and does not provide conclusive results [27; 38].

Our work add to this literature by providing empirical evidence on the effect of platforms Price Matching Guarantees policies on daily consumer electronics prices observed on US online market. We have focused on the NewEgg platform that exclusively sells consumer electronics products and implements PMG policies that turn on and off over time (blinking PMG). Given that our identification strategy is based on a comparison of price levels before and after the policy shutdown, we excluded platforms that never stop offering PMG (like Target).

In particular, we apply a Diff-in-Diff (DiD) approach where we consider as the treated sample the pool of NewEgg products interested by PMG policies. Differently from standard practices in studies adopting a DiD approach, we build the control sample with price data for the same products observed on a different platform, namely Amazon, that never offers PMG to customers. Furthermore, in order to ensure that our counterfactual sample is less likely to be influenced by the PMG policy adopted by NewEgg, we have considered data from the Amazon UK platform, instead of Amazon US. Indeed, price observed on Amazon US might not be completely independent from the policy under scrutiny, because of price tracking practices frequently adopted by platforms.

Estimates provide evidence in favor of an average price reduction of about 3.9% after the interruption of the PMG policy. However, in order to have a more detailed picture of the issue, we investigate if such effect is heterogeneous, depending on products characteristics. In particular, we focus on products features that might affect the outcome of PMG policies and that can be recovered exclusively on online markets. Platform data allow us to obtain information from User Generated Contents (UGC), like product popularity, product quality and online search visibility (Google Search Rank); indeed, we believe that these product characteristics might provide indirect information on consumers heterogeneity. Estimates conducted on specific sub-samples show that when the PMG is interrupted, low quality (and low search rank) products experience a price increase of about 3.4%, while for high quality (and high visibility) products a price reduction of about 3.7% is observed. These findings are in line with the lack of unambiguous predictions of the theoretical literature.

The anti-competitive effects of PMG observed for high quality (visibility) products has been predicted by theoretical models where such polices make collusion more likely [7; 20; 34]. These findings can be also explained by the theoretical predictions of a class of models, like Corts [14] and Nalca et al. [30], where PMG is a tool for discriminating customers according to their sensitiveness to price and products quality. These models also explain our results obtained for products characterized by low visibility (quality). Indeed, the willingness to engage in search activity could indirectly identify those customers whose demand is more rigid, as argued by the search literature [17].

The analysis conducted in this study enriches the literature on the price effects of PMG by using very detailed platform daily price data for a specific market (consumer electronics) where such policies are very common. First, our work overcomes previous research (see Zhuo 38) by using real-time data instead of historical information. This makes the use of price tracking websites and algorithms that extract data from price charts outdated.
Second, the DiD identification strategy adopted is based on the construction of a control group with a novel approach; finally , products characteristics based on Users Generated Contents (UGC) are employed for the first time in order to study possible heterogeneous effects of such policies.

The rest of the paper is organised as follows. In the next section we briefly discuss related literature and in Section 3 we accurately describe the data extraction process and the variables used in the empirical analysis. Section 4 explains our identification strategy and in Section 5 we discuss estimates results and robustness analysis. Section 6 concludes.

2 Literature Review

The theoretical literature has analysed possible impacts of Price Matching Guarantees (PMG) on different market outcomes, since such commercial policies might affect the behaviour of firms and consumers in different ways.
The most common prediction of the theoretical models is that price matching guarantees hamper competition by keeping prices high and sustaining collusive practices; moreover, some authors suggest that they might be tools for realising price discrimination or signalling cheap prices.

Hay [20], Salop [34] and Belton [5] have first suggested that price matching guarantees can sustain collusion in oligopoly models; they highlight that such clauses might be considered as threats to punishment for firms that lower cartel prices, thus reducing firms incentive to deviate from the agreement. They argue that, if all competitors in the market adopt a PMG policy, none of them has the incentive to lower its price and the latter tends to the monopolistic level. Moreover, they agree on the fact that the adoption of such policies increase the stability of the cartel, as any price cut must be refunded to the consumer, so that the policy generates a credible penalty system.
Several other papers support the pro-collusive argument by extending the basic oligopolistic setting (see among others Doyle [15], Logan and Lutter [25] and Baye and Kovenock [4]), while other authors explore the impact of PMG extending the analysis in dynamic, multi-stage and Hotelling frameworks (see among others Chen [9], Lu and Wright [26], Hviid and Shaffer [23], Pollak et al. [33], Constantinou and Bernhardt [13]). Cabral et al. [7] suggest that a PMG can be a collusion enacting practice. In the model two firms alternate over time in setting prices; given that starting a collusion process implies several risks, like for example antitrust penalties, firms include collusion costs in their decisions. The main prediction of the model is that the probability of tacit

collusion rises when the policy is in place.
In studies reviewed so far, it is implicitly assumed that customers automatically claim the price guarantee whenever they find a price differential: indeed, this is not always the case, because of lack of information or because there are small costs for the customer to activate a guarantee, the so called “hassle costs”. Hviid and Shaffer [22] highlight that the presence of hassle costs undermines possible anti-competitive effects of PMG, but do not completely cancel them. Precisely, with symmetric firms PMG are unable to support any price increase in presence of hassle costs. Indeed, each firm will be interested in lowering price levels by an amount that is marginally smaller than these costs, so that buyers are attracted from cheaper firms and do not activate the guarantee. Otherwise, with asymmetric firms, a rise in prices might be supported, but not at the monopolistic level. Moreover, their model can explain why universal adoption is not a realistic assumption of previous studies.

Some other models explore the possibility that sellers use PMG policies as a price discrimination tool. If customers are different in terms of some subjective characteristic, like information on prices and guarantee terms, degree of loyalty to a specific retailer or level of hassle costs in requesting the refund, firms could use the price guarantee to discriminate between different groups of consumers. Png and Hirshleifer [32], Belton [5] and Corts [14] first suggested duopoly models where firms discriminate between different consumers groups, namely “unsophisticated” customers and “sophisticated” ones. Consumer segmentation and PMG allow firms to set higher prices for unsophisticated consumers, while sophisticated ones benefit from the lowest price guaranteed by the policy. The main intuition from this strand of literature is that price discrimination might at least benefit some customers with actually lower prices.[58]
Finally Moorthy and Winter [28] suggest that PMG might be a credible signal of low prices, if low cost firms adopt the policy and (high cost) competitors can not match the policy. Similarly, Jain and Srivastava [24] develop a theoretical model that identifies the conditions under which PMG might lead to lower market prices.[59] In the presence of informed and uninformed consumers (about prices and store characteristics) and of different kind of stores (in terms of size, service quality ans so on), only stores with low prices offer price-matching policies.[60]

Despite the theoretical literature is rich and analyses several aspects of price matching policies, the empirical evidence is scant and does not provide conclusive results. Some studies focus on specific markets, like tyre or gasoline, while others analyze retailing prices from supermarkets, grocery stores or online platform markets.

Analysing daily price quotes from the tyre industry advertisements, on 61 US Sunday newspapers observed for three months in 1996, Arbatskaya et al.

[58] Similar results can be found also in Edlin [16] and Nalca et al. [30].

[59] Authors have realized two experiments to analyze the effect of PMG policies on prices consumer perceptions and have shown that consumers did expect lower prices from PMG.

[60] Similar results can be found also in Moorthy and Zhang [29].

[1], through a Feasible Generalised Least Square approach, find weak evidence of anti-competitive effects of PMG and show that an increase in the number of firms implementing the policy leads to a 10% increase in prices.[61] Cabral et al. [7], focus instead on daily pricing policies adopted by the Shell network of gas stations in Germany in 2015.[62] Leveraging on gas stations localization and consumers demographics as sources of identification, they suggest that PMG can be a collusion enacting policy. Gas station prices have been analyzed also by Byrne and De Roos [6] for Australia by means of a detailed 15 years time series dataset. Authors argue that the majority of gas stations prices follow a weekly cycle and that dominant firms can use PMG to coordinate market prices and reduce price competition. Similar results can be found in Chilet [10], who analyses pricing policies of three big retail pharmacy chains in Chile, observed over the period 2006-2008. The author follows an identification strategy based on the estimation of a demand model, in which quantity sold is a function of the differences between own prices with the competitors ones, around the time period where collusive price increases occurred.

Hess and Gerstner [21] analyse the effect of PMG on prices by collecting weekly data of 114 goods sold in several US supermarkets and grocery stores, from 1984 to 1986. Authors, by means of a time series analysis, provide evidence in favor of higher prices of about 1-2% when the guarantee is introduced. Moorthy and Winter [28] argue instead that the adoption or non-adoption of the PMG might be interpreted as a way to signal the seller service–price profile.[63] Authors analyse data for several product categories from 46 Canadian retailers observed in 2002. They assume the existence of informed and uninformed consumers and show that PMG might be a tool to signal low prices to uninformed consumers. In particular, they find that PMG are mainly adopted by low cost/low service chain stores. Similar results can be found in Chung et al. [11] for three leading hypermarkets in Korea. Finally, Zhuo [38] focuses on online platforms and collect US price data from online price trackers for 150 products offered on Amazon in 2012. The author observes prices during and after the implementation of PMG policies by two big-box stores (Target and Best Buy) targeted specifically on Amazon prices; by applying DiD and RDD methods, the author suggests that prices increase by about six percentage points during the period of validity of the policy;[64] moreover, the analysis highlights an heterogeneous impact of PMG, with larger price increases for initially lower-priced goods.[65]

[61] The same authors in Arbatskaya et al. [2] confirm their results by analysing the same data with a different approach.

[62] See also Wilhelm [36].

[63] The authors refer to the retailer service-price profile as to any sellers characteristic that might induce customers to choose one seller over another one, like better sales assistance and customer care, a clear Web site, personalised delivery and selling services.

[64] Similar results can be found also in Wu et al. [37], Haruvy and Leszczyc [19].

[65] Some other authors analyse the impact of price-beating guarantees, that are less widespread policies with similar terms as price matching ones (in price beating guarantees refund exceeds the price difference). Studies that refer to these policies argue that, with respect to price matching guarantees, they might be serving different purposes in prac-

3 Data

3.1 Data Extraction

In order to study the impact of PMG on prices, we focus on the online consumer electronics market, since it is one of the most widespread sector on online retailing and is often interested by such pricing policies. In particular, electronic products are search goods, whose quality can be evaluated before the purchase: the advent of online markets has made this process much cheaper and faster and is most likely to affect the impact of such policies, whose outcome depends, among other factors, by the level of search and hassle costs. Moreover, electronic goods are barely affected by seasonal effects, so that prices signals are more stable over time and show low price differentials across countries [18; 35]. These characteristics allow us to improve the identification strategy through the construction of a more refined control group (see the next section).

Among different online retailing platforms we choose to focus on NewEgg, a leading online US retailer of consumer electronics products, that implements a so-called blinking PMG, i.e. a price guarantee that turns on and off over time on selected items. Given that our identification strategy is based on the comparison of prices before and after a policy shutdown, we do not consider platforms that apply PMG to wide groups of products continuously over time (i.e. Target, among others). In particular, NewEgg communicates the period of validity of the price guarantee by means of a label that appears on the specific product online page; the customer who discovers the PMG badge has 14 calendar days of time to find the same title at a lower price from US competitors belonging to a declared list.[66] PMG policies are often repeated over time on the same products without any notice, so that consumers looking for deals have to exert an higher effort in the search process.

In order to build the sample we have identified 100 NewEgg products interested by PMG on May 10^{th} 2018. For such products we have collected price data and the presence of the NewEgg PMG badge until 31^{st} October 2018 (174 days and 9028 observations). We identify as the treatment of interest the interruption of the PMG policy, so that prices observed on the NewEgg platform represent the treated sample. The control sample has been built by recovering price data for the same products observed on NewEgg but sold on the Amazon UK platform, that never offers PMG policies.[67] This reduces the number of observed products, so that the final sample includes 29 products belonging to 19 sub-categories (computer hardware, tablet and computers, mobile phones, printers and scanners, PC accessories, speakers for domotics, screens and au-

tice and likely be effective in enhancing competition. Experimental literature also focuses on the effect of price matching and price-beating guarantees: however, experimental results lack the complexity of real interactions between sellers and consumers.

[66] With title we refer to a product with the same brand and model number. NewEgg, after checking the validity of the claim, sends a Customer Care Card to refund the price difference (Source: https://kb.newegg.com/knowledge-base/price-match-guarantee/).

[67] See the next section for a rationale on this choice.

dio devices). In the Appendix we provide a detailed list of selected products (Tables 30 and 31).
It is worth noting that both Amazon and NewEgg operate either as online retailers or markeplaces for third-party sellers who pay fees and royalties to access to customer base. In such marketplace, online platforms often acts only as a payment intermediary and goods are kept in the third-party sellers inventory. Thus, in order to build a valid control group, we have excluded data on products sold by third-party sellers on both Amazon and NewEgg platforms.

The retrieving of sample data has been a challenging task. Given the absence of ready-made and easy-to-use repositories on price data, we have developed an ad-hoc scraping program (in Python language) able to protect the scraping process from unpredictable changes of the page and capable to recover the data without stressing the site, thus limiting the risk of interruptions due to firewalls.[68] In particular, the scraping process has been supported by several alert tools signalling periodical changes of the internal page structure.[69]

The process of data collection has required the daily implementation of these main steps:

1. Sign up for subscription to Amazon Web Service (AWS) cloud, in order to use virtual servers in which to install and launch the program;[70]

2. Accept the norms and terms of use of the platform site, in order to be compliant to the server navigation policies;

3. Launch the daily loop process, in order to navigate among product pages, select the field tags, get the data and save on a server disk. Each scraping session runs about 20 minutes every day.

In addition to products daily prices retrieved on both platforms, we also collect several product characteristics available exclusively on online sales platforms. In particular, we recover some User Generated Content (UGC), like the absolute number of reviews, received by the specific product under consideration and the most popular one in the same subcategory, as the product rating; moreover, from Google we perform and collect a product search rank.
The absolute number of reviews is a dynamic information which represents a sort of popularity index, since it is proportional to the product market diffusion.[71] We also calculate the relative number of reviews as the ratio between the number of reviews of each product and the amount of reviews received from the most popular product in the same subcategory.[72] This normalized index,

[68]A typical problem is to intercept daily changes of web pages not only about prices, but also concerning other dynamic contents, such as the number of customer reviews, the average rating and so forth. Code available from authors upon request.

[69]Indeed, platforms frequently change the deep structure of the page, in a not visible way by the human reader but in a way that affects the program code and the scraping process.

[70]AWS is a comprehensive, evolving cloud computing platform provided by Amazon.

[71]In online commerce, product reviews are used by retailing platforms to give consumers an opportunity to comment on products they have purchased, right on the product page.

[72]See Table 30 in Appendix for details.

that ranges from zero to one, shows the relative popularity of the product with respect to the other items of the same sub-category. We also collect data on product ratings (stars) provided by consumers. We consider the number of stars gained by each product, ranging from zero (low quality) to five (high quality), as a proxy of product quality. Finally, we develop a search index as a proxy of the time spent on search engines to discover the page of a certain product. More precisely, the search index represents the probability to find the product in the first ranked positions of Google results.[73]

It is worth noting that, although the products analysed are sold by Amazon and NewEgg in different countries, information on some of the considered UGC (e.g. rating) maintain their consistency. This property is typical of consumer electronics goods that have a standardized nature. Concerning the search index, we adopt a country specific value by launching the Google search engine with specific country settings (UK and US).

3.2 Descriptive Statistics

Our sample consists of 9028 daily price observations (174 days) for 29 products observed on NewEgg and Amazon UK platforms, from 10 May 2018 until 31 October 2018. Table 19 shows the mean and the standard deviations of prices and selected product characteristics for the overall sample and for the treated and control group ones. Prices show a large variability, being the average for the overall sample $240.43 and the standard deviation $283.53. By comparing average values observed over the two platforms, it emerges that both prices and UGC display similar values, thus confirming what has been observed by the previous literature on the low dispersion of consumer electronics prices across countries [35]; moreover, such similarities support our approach for the choice of the control sample. As Table 20 shows, in the case of the treated sample (NewEgg) the average price during the policy validity period (before treatment) is about $18 higher with respect to the post implementation period.[74]

In order to investigate the issue of heterogeneity in the effect of PMG policies on prices, we distinguish products according to products characteristics recovered from UGC. In particular we classify products depending on their quality and visibility, measured trough UGC as explained in the previous section. Given that quality assessment by consumers is highly correlated to products visibility, in Table 21 we show some descriptive statistics for products classified according to such characteristics.[75]
Again, data show that products characteristics stemming from UGC are quite similar across countries/platforms.

[73] The ranking position of an item is retrieved launching the Google query composed by the sentence: “the name of product” AND “the name of platform”. The resulting position is then normalized, mapping in the probability range [0,1].

[74] We remember that our treatment is the policy shutdown.

[75] High quality products are those characterized by a rating higher than 4/5, while high visibility ones are those endowed of a search rank index greater than 0.8.

Table 19: Summary Statistics. Treated and Control Samples.

Variables	Full Sample	Amazon UK	NewEgg
Provider Price ($)	240.43	227.72	253.15
	(283.53)	(262.74)	(302.39)
Product Popularity (0-1)	0.23	0.26	0.20
	(0.27)	(0.30)	(0.23)
Search Rank (0-1)	0.75	0.85	0.64
	(0.30)	(0.17)	(0.36)
Rating (0-5 stars)	4.14	4.14	4.15
	(0.68)	(0.48)	(0.83)

Table 20: Summary Statistics. NewEgg. Pre and Post Treatment.

Variables	Pre Treatment	Post Treatment
Provider Price ($)	231.22	209.63
	(329.01)	(235.21)
Product Popularity (0-1)	0.22	0.19
	(0.25)	(0.23)
Search Rank (0-1)	0.77	0.71
	(0.26)	(0.31)
Rating (0-5 stars)	4.16	4.06
	(0.98)	(1.05)

Notes: The pre-treatment period is the policy implementation period.

Table 21: Summary Statistics. Sub-Samples.

Variables	Low Quality Low Search Rank		High Quality High Search Rank	
	NewEgg	Amazon.uk	NewEgg	Amazon.uk
Provider Price ($)	206.02	95.81	221.06	161.10
	(143.92)	(35.11)	(354.69)	(163.90)
Product Popularity (0-1)	0.01	0.01	0.28	0.29
	(0.01)	(0.01)	(0.25)	(0.33)
Search Rank (0-1)	0.19	0.38	0.90	0.91
	(0.27)	(0.38)	(0.04)	(0.05)
Rating (0-5 stars)	2.93	3.19	4.40	4.45
	(0.71)	(0.32)	(0.49)	(0.20)

Notes: For high quality products we mean those with ratings higher than 4. For high visibility products we mean those with a normalized search index higher than 0,8.

As far as the PMG policy is concerned, NewEgg adopts a blinking strategy, so that the policy is applied in a non continuous way, often to the same products. Table 22 shows the total number of days of treatment (absence of PMG) and the average number of treatments occurred in each sample. This latter information suggests that, on average, the policy is applied to each product twice during the sample period (174 days) and such frequency does not seem to be correlated to products quality and visibility. Indeed, since prices are highly correlated to quality, we can reasonably assume that there is not selection into treatment associated to products price or quality (visibility), so that the assumption of random assignment required by the identification strategy seems reasonably fulfilled. On the other side, it seems that, for low quality (and visibility) products, the policy implementation period is longer.

Table 22: Summary Statistics. PMG.

Variables	Full Sample	Low Quality	High Quality
		Low Search Rank	High Search Rank
Treatment Duration (days)	38.25	58.13	29.30
	(39.94)	(43.37)	(33.54)
Number of Treatments	2.38	1.81	2.78
	(1.75)	(0.82)	(2.06)

Notes: Treatment duration is the average number of days without PMG. The sample period includes 174 days.

Another important issue is related to the representativeness of our sample. Figure 6 represents the distribution of products by price classes (10). The graph shows that 22 products out of 29 belong to the first two price deciles, with price ranging between 0$ and 240$. This picture closely matches a typical distribution observed in consumer electronics [12], often characterized by a large amount of low cost accessories and few luxury goods. Furthermore, calculating the log-price distribution (Figure 7) and mapping the integer part of this value on the x-axis, we obtain a septile-partition. By plotting the distribution of products by log-price classes we obtain a distribution that resembles the Normal one. Such result is in line with those obtained by Coad [12].

Figure 6: Products Distribution by Price Classes.

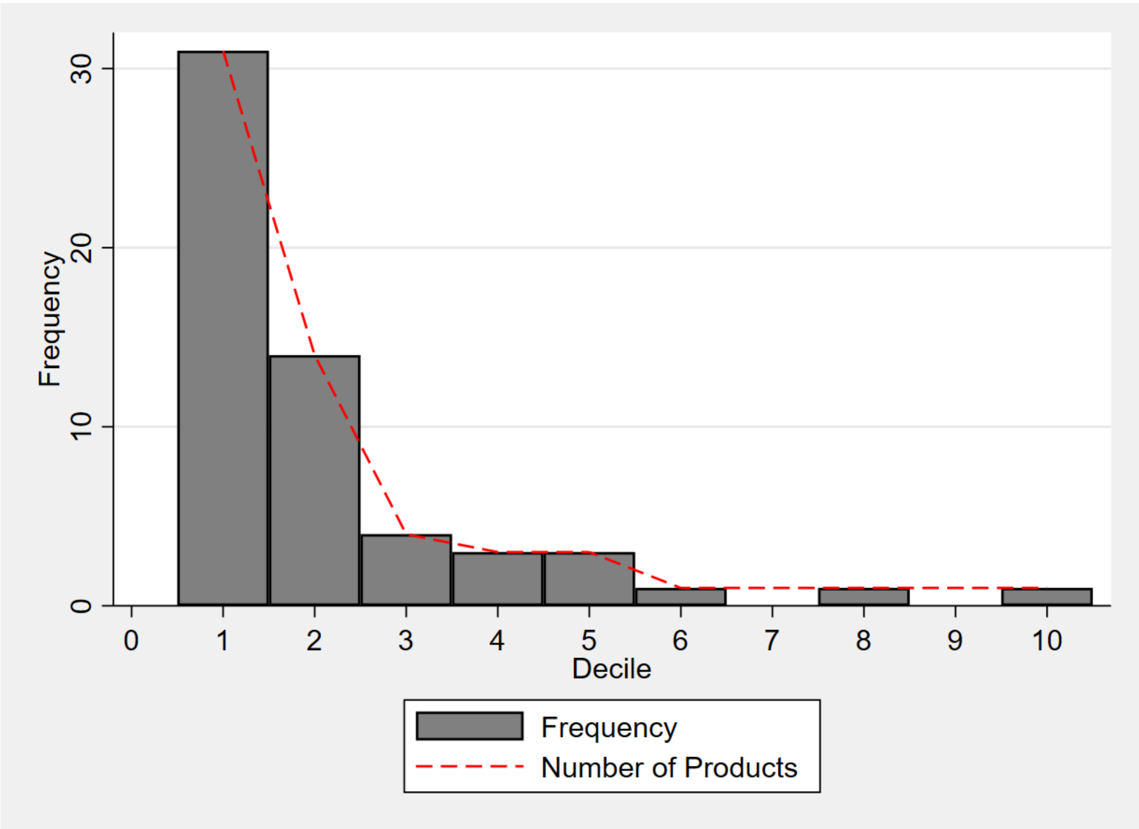
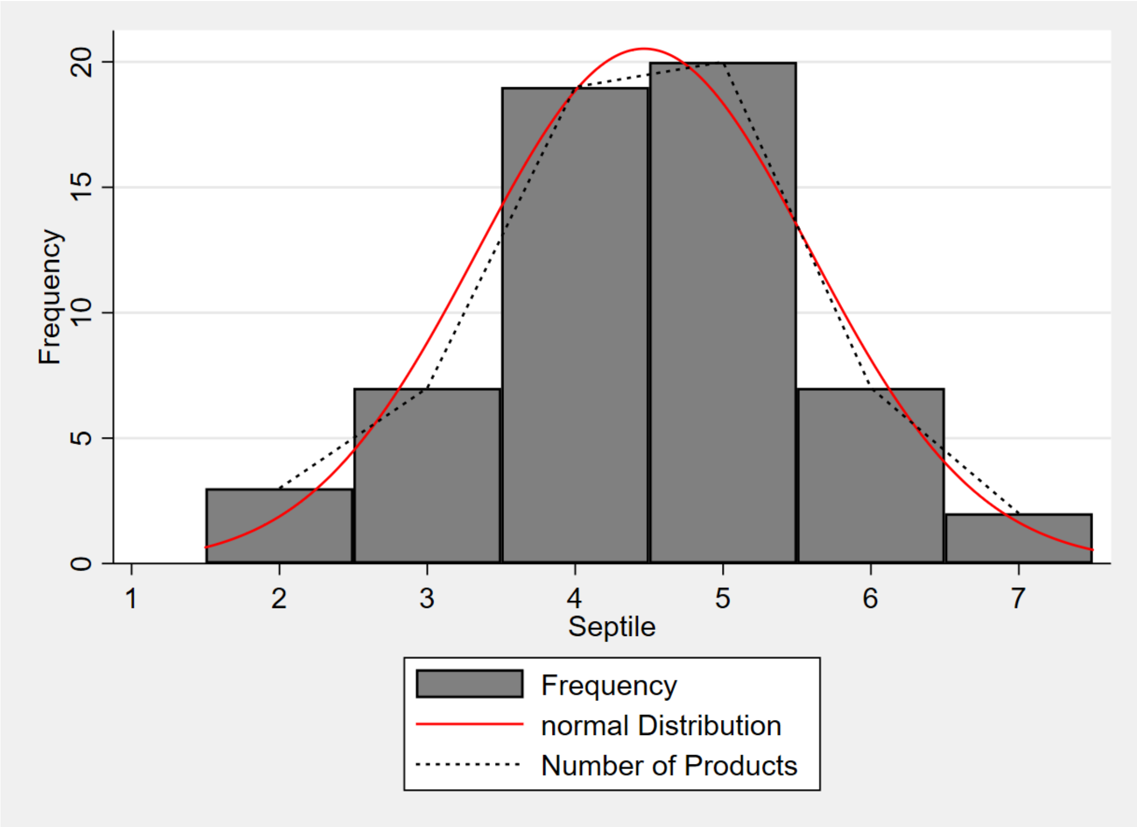


Figure 7: Products Distribution by Log-Price Classes.



4 Identification Strategy

We identify the causal effect of price matching guarantees on price levels, by comparing prices before and after the policy shutdown for a sample of products sold by NewEgg (the treatment group), to the prices average change for the same products sold by Amazon UK (the control group). Indeed, and crucially for our identification strategy, PMG implemented by NewEgg only affects products that are sold in US, thereby naturally creating a treatment and a control group; the same products sold by Amazon UK (that never offers price warranties) are less likely to be affected by the policy and well represents a counterfactual sample mimicking what would have happened to prices of treated products in the absence of PMG. This framework provides a quasi-natural experiment that allows us to study the causal impact of PMG on prices through a Diff-in-Diff research design.

This identification approach requires the estimation of the following panel FE model:[76]

$$\log Price_{i,l,t} = \alpha_{i,l} + \gamma\big(T_{i,l,t} * P_{i,t}\big) + X^{T}_{i,l,t}\beta + \mu_{i,l} + \tau_t + \epsilon_{i,l,t} \qquad (15)$$

[76]In a Diff-in-Diff context, a classic model would be built like $Y = \alpha + \beta_1(Treated) + \beta_2(Post) + \beta_3(Treated * Post) + \varepsilon$. In all models we exclude $Treated$ and $Post$ terms, since they are multicollinear with time and product fixed effects.

The dependent variable, $Price_{i,l,t}$ represents the price (natural logarithm) of good i, on platform l, at time t; $T_{i,l,t}$ denotes a binary variable equal to 1 for treated goods; $P_{i,t}$ is a binary variable that is equal to 1 for any day since the policy shutdown and zero otherwise and $\epsilon_{i,l,t}$ is an error term. The model includes a full set of daily time dummies, τ_t, accounting for unobserved time-varying determinants of prices that are common to all goods. Product fixed effects, $\mu_{i,l}$, control for any time invariant unobserved heterogeneity at the product and platform level, that could be correlated with the included regressors and that could also drive prices. Moreover, the presence of individual (product) fixed effects in the Diff-in-Diff research design rises the degree of comparability of treatment and control groups.

We include a set of covariates, $X_{i,l,t}$ in Equation (15), in order to control for products characteristics derived by UGC that might affect the outcome of the PMG policies. The γ coefficient associated to the interaction term $(T_{i,l,t} * P_{i,t})$ represents the DiD estimate of the effect of PMG shutdown on treated products prices and it measures the average price differential between the treated and the control group.

We also explore the issue of heterogeneity in the effect of PMG policies on prices. Indeed, as discussed in the literature review section, most of the predictions of theoretical models on the price effects of PMG policies rely on assumptions related to the presence of heterogeneous consumers. By distinguishing products according to consumers quality assessment, we indirectly assume that consumers are heterogeneous in terms of their preferences towards quality and their availability to pay a price premium for that. Indeed, for high quality goods the price elasticity of demand is usually assumed to be lower than the price elasticity for low profile goods. We further classify products according to their visibility, as measured by the search index described above. We believe that the time spent for finding a product indirectly selects consumers according to their willingness to engage in search activity and that such availability is directly correlated to their price sensitiveness.

Based on the above reasoning, we estimate Equation (15) on different subsamples built according to product quality and visibility indices. In particular, we analyse separately high (low) quality products, namely products characterized by rating greater (lower) than 4/5, and products characterised by high (low) visibility in terms of Google search rank, namely products whose search index is greater (lower) than 0,8. Moreover, given that products quality and visibility resulted to be highly correlated, we split the sample according to both characteristics. As discussed in the Data Section, such products characteristics do not affect the probability of being treated.

The heterogeneity issue is also investigated with a different approach by estimating a Triple Difference regression (DDD) on the full sample. In particular, we estimate the following model:

$$\log Price_{i,l,t} = \alpha_{i,l} + \varphi\big(T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}\big) + X^T_{i,l,t}\beta + \mu_{i,l} + \tau_t + \omega_{i,l,t} \quad (16)$$

Equation (16) includes an additional component in the interaction term, $HRHV_{i,l,t}$, i.e. a dummy variable assuming value 1 for high quality and high visibility products. The coefficient φ of the triple interaction measures the average treatment effect of PMG on prices for high quality (visibility) products.

All specifications are estimated by OLS and Standard Errors are robustly estimated. Moreover, following Cameron and Miller [8], we compute bootstrapped standard errors with a cluster structure (at product level) and all results are confirmed. Finally, we conduct an extensive robustness analysis through different falsification and placebo tests (see the Robustness Analysis Section).

5 Empirical Results

5.1 Main Results

In Table 23 and 24 we show DiD estimates; in particular, we first report results obtained by estimating Equation (15) without including control variables (Table 23, column (1)), while Table 24 (column (1)) provides estimates obtained after including all control variables. DiD estimates suggest that the PMG shutdown triggers a significant reduction of price levels of about 3.9%. Rather comfortingly, the inclusion of control variables into model (15) does not significantly affect the result. These findings suggest that, on average, the adoption of PMG has an anti-competitive effect on prices since, after the policy validity period, they show a substantial reduction. These results are consistent with those obtained by Zhuo [38] on a large sample of products observed on the Amazon platform in 2012. However, we follow a rather different identification strategy. While Zhuo [38] focuses on price changes observed on the non-adopting platform, before and after the implementation of PMG by competitors, we focus on price changes observed on the adopting platform. Moreover, we innovatively build the control sample with platform price data for the same treated products but observed in another country (UK).

To explore whether product properties affect the impact of PMG on prices, we split the sample according to different classes of product quality and visibility and we re-estimate Equation (15). Columns from (2) to (5) in Tables 23 and 24 show results of this disaggregated analysis. Estimates indicate that a policy repeal produces a price reduction for both low and high quality products; however, the estimated coefficients for the low quality sample are not statistically different from zero, while those for high quality products indicate a statistically significant price reduction of about 2.5%. When we split the sample according to values assumed by the search index, results suggest that, when the PMG is interrupted, products characterised by a low search rank experience a price increase of roughly 2,4%, while for high visibility products prices decreases of about 5,3%. These findings support the hypothesis that, in online consumer electronics market, PMG policies harm competition for high visible products by keeping prices high, while for low visible products, such policies have a pro competitive effect on prices.

Indeed, as highlighted in the data section, quality and visibility are highly correlated in our sample. Hence, we estimate Equation (15) after splitting the

sample according to both product properties.
Results shown in column (6) and (7) of Table 23 suggest that the PMG shutdown triggers a reduction of prices for high quality and high visibility products (3,7%), while prices of low quality and low visibility ones raise of about 3,4%. These findings are confirmed when we include control variables into the model (Table 24, columns (6) and (7)) and when we analyse heterogeneous effects of PMG by means of a Triple Difference regression approach, as shown in columns (1) and (2) of Table 25.

Table 23: DiD Estimates of the Impact of PMG on Prices.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Products Prices (log)	FULL SAMPLE	L. RATING	H. RATING	L. VISIBILITY	H. VISIBILITY	LR-LV	HR-HV
$T_{i,l,t} * P_{i,t}$	-0.0401***	-0.0064	-0.0250***	0.0242***	-0.0543***	0.0331**	-0.0381***
	(0.00628)	(0.01300)	(0.00688)	(0.00769)	(0.00795)	(0.01370)	(0.00864)
Observations	9,028	2,896	6,132	2,295	6,733	994	4,864
R-squared	0.986	0.985	0.986	0.990	0.984	0.983	0.983
Controls	NO	NO	NO	NO	NO	NO	NO
Product Dummies	YES	YES	YES	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES	YES	YES	YES
F Test	0.000	0.623	0.000	0.002	0.000	0.016	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. High quality products have ratings higher than 4. High visibility products have a normalized search index higher than 0,8. LR-LV are low rating and low search index products, HR-HV are high rating and high search rank products. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 24: DiD Estimates of the Impact of PMG on Prices.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Products Prices (log)	FULL SAMPLE	L. RATING	H. RATING	L. VISIBILITY	H. VISIBILITY	LR-LV	HR-HV
$T_{i,l,t} * P_{i,t}$	-0.0424***	-0.0108	-0.0270***	0.0322***	-0.0577***	0.0532***	-0.0398***
	(0.00629)	(0.01510)	(0.00693)	(0.00786)	(0.00799)	(0.01310)	(0.00879)
Observations	9,028	2,896	6,132	2,295	6,733	994	4,864
R-squared	0.986	0.986	0.986	0.992	0.985	0.989	0.983
Controls	YES	YES	YES	YES	YES	YES	YES
Product Dummies	YES	YES	YES	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES	YES	YES	YES
F Test	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. High quality products have ratings higher than 4. High visibility products have a normalized search index higher than 0,8. LR-LV are low rating and low search index products, HR-HV are high rating and high search rank products. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Our empirical findings can be explained by the main predictions of theoretical models analysing the impact of PMG on prices and competition.
The anti-competitive effect of PMG observed for high quality (visibility) products has been predicted by theoretical models where such polices make collusion more likely [7; 20; 34]. These findings can be also explained by the theoretical predictions of a class of models, like Corts [14] and Nalca et al. [30], where PMG is a tool for discriminating customers according to their sensitiveness to price and products quality. These models also explain our results obtained for products characterized by low visibility (quality).
Indeed, most of the predictions of theoretical models on the price effects of PMG policies rely on assumptions related to the presence of heterogeneous consumers. By classifying products on the base of consumers quality assessment, we indirectly assume that consumers are heterogeneous in terms of their preferences towards quality and their availability to pay a price premium for that. Similarly, the time spent for finding a product can indirectly select consumers according to their willingness to engage in search activity and it is

Table 25: DDD Estimates of the Impact of PMG on Prices.

Products Prices (log)	(1) DDD	(2) DDD
$T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}$	-0.0537***	-0.0556***
	(0.00808)	(0.00810)
Observations	9,028	9,028
R-squared	0.986	0.986
Controls	NO	YES
Product Dummies	YES	YES
Time Dummies	YES	YES
F Test	0.000	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. $HRHV_{i,l,t}$ is a dummy equal to 1 for high quality and high visibility products. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

reasonable to argue that such availability is directly correlated to price sensitiveness.

5.2 Robustness Analysis

In this section, we discuss empirical results obtained by conducting an in-depth robustness analysis of our results.

The first issue that we tackle is the possibility that the effects of the treatment speed up, stabilize, or mean revert over time. In order to explore this issue, we estimate a specification of Eq. (15) that includes lags à la Autor [3] and takes on the following form:

$$\log Price_{i,l,t} = \alpha_{i,l} + \sum_{j=0}^{5+} \gamma_j \big(T_{i,l,t} * P_{i,t+j}\big) + X_{i,l,t}^T \beta + \mu_{i,l} + \tau_t + \epsilon_{i,l,t} \tag{17}$$

where $P_{i,t+j}$ assumes the value of 1 in day $t + j$, and 0 otherwise. Specification (17) allows the PMG repeals to generate different effects over time. In order to lower the number of parameters of the model, we estimate the effect of a PMG shutdown from the implementation day ($j = 0$) until five days later and onward.

According to results shown in Table 26, coefficients related to lagged variables are always negative and statistically significant for the full sample. However, point estimates suggest that the impact of the treatment reaches its maximum after one day and starts decreasing afterwards. Figure 8 graphically shows parameter estimates patterns.

Another important issue in a DiD research design is the presence of pretreatment common trends for treated and control samples. This assumption is indeed fundamental for the validity of the counterfactual policy evaluation analysis.

Table 26: DiD Estimates of the Impact of PMG on Prices with lags à la Autor (2003).

Products Prices (log)	(1) DiD	(2) DiD
$T_{i,l,t} * P_{i,t+0}$	-0.0500***	-0.0508**
	(0.01610)	(0.01610)
$T_{i,l,t} * P_{i,t+1}$	-0.0566***	-0.0577**
	(0.01580)	(0.01570)
$T_{i,l,t} * P_{i,t+2}$	-0.0558***	-0.0566**
	(0.01960)	(0.01960)
$T_{i,l,t} * P_{i,t+3}$	-0.0530***	-0.0539**
	(0.01990)	(0.01990)
$T_{i,l,t} * P_{i,t+4}$	-0.0529***	-0.0531**
	(0.02030)	(0.02030)
$T_{i,l,t} * P_{i,t+5+}$	-0.0368***	-0.0395**
	(0.00646)	(0.00648)
Observations	9,028	9,028
R-squared	0.986	0.986
Controls	NO	YES
Product Dummies	YES	YES
Time Dummies	YES	YES
F Test	0.000	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Figure 8: DiD Estimates of the Impact of PMG on Prices (Full Sample, with Controls) with lags à la Autor (2003).

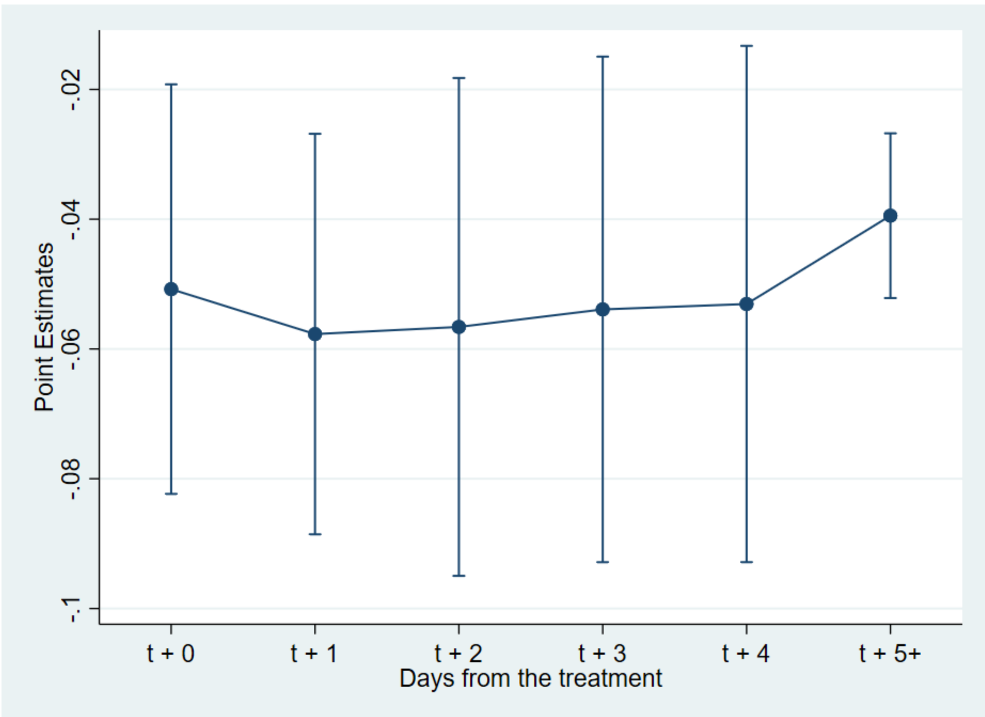
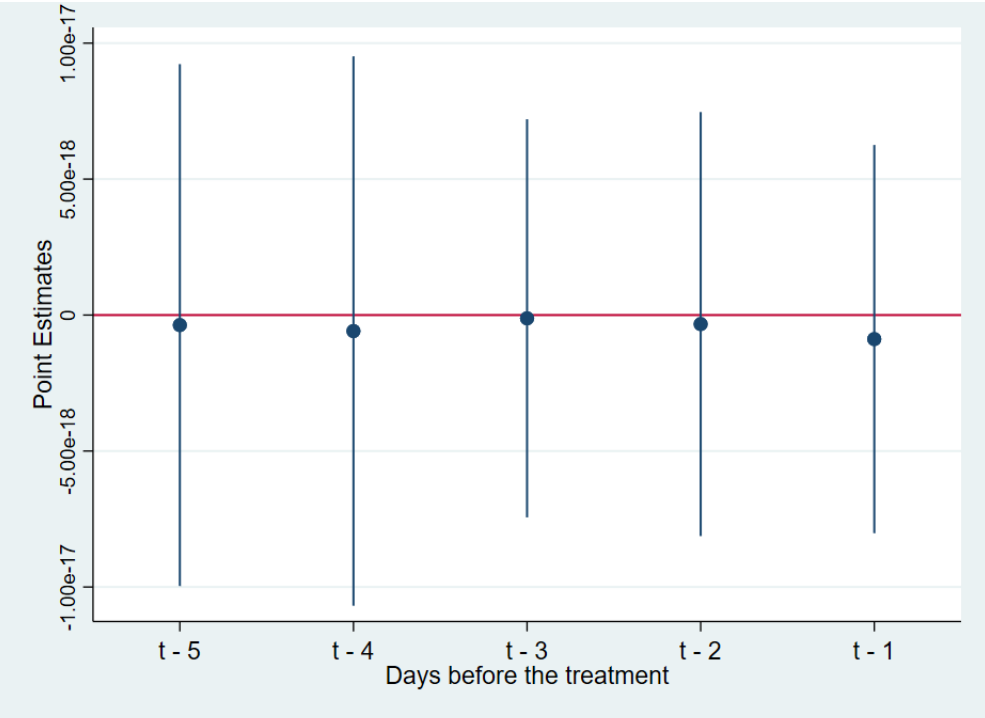


Figure 9: Price Differentials Between Treated and Control Groups Before the PMG Shutdown.



In order to explore this issue, we show in Figure 9 point estimates values and the relative confidence intervals of the difference in the level of prices between treated and control products from five days before the treatment to the day of the policy shutdown.[77] Plotted point estimates suggest that price levels for the treated platform do not seem to be significantly different from prices of the control platform before the treatment. This result provides evidence in favor of the validity of parallel trends assumption for our samples.

In order to further analyse this issue, we follow Autor [3] and we estimate Eq. (17) after including some leads of the treatment interaction variable:

$$\log Price_{i,l,t} = \alpha_{i,l} + \sum_{j=-1}^{-5} \gamma_j \big(T_{i,l,t} * P_{i,t+j}\big) + X_{i,l,t}^T \beta + \mu_{i,l} + \tau_t + \epsilon_{i,l,t} \qquad (18)$$

If leads coefficients turn out to be statistically significant, there may be anticipatory effects and a failure in the parallel trend assumption. According to Table 27 and Figure 10, estimated coefficients of the anticipatory effects are not statistically significant, thus providing further evidence in favor of the existence of a parallel trend between treatment and control sample.

In order to extend our robustness analysis, we implement a complete set of placebo tests. We first estimate our baseline and DDD specifications by intro-

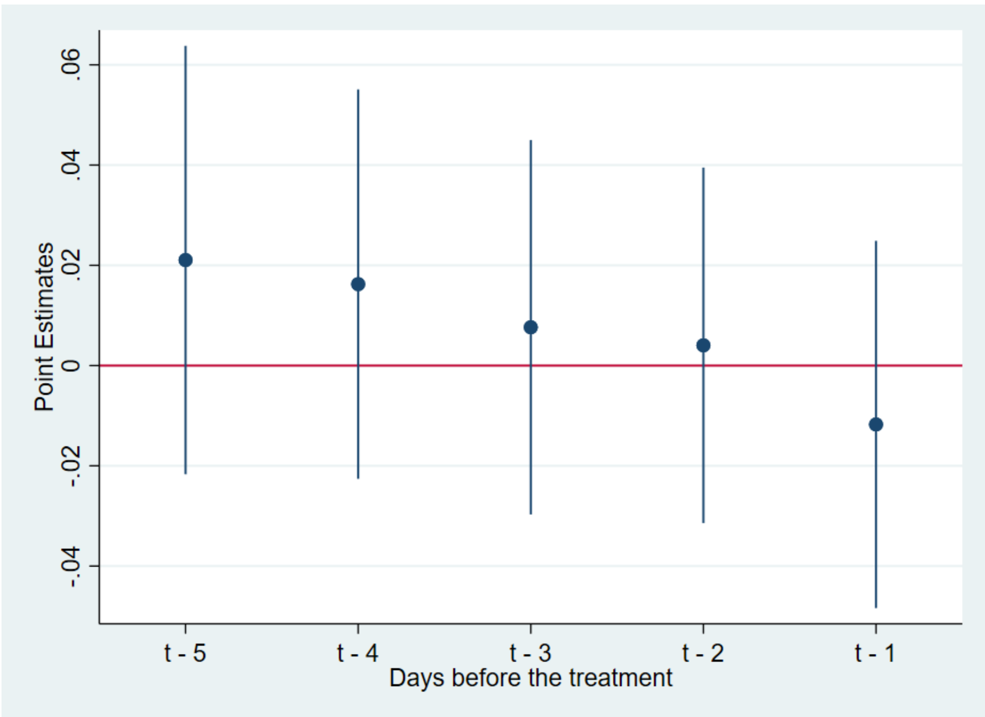
[77]In order to obtain these values, we estimate a panel model where we regress average daily price differences between the two samples on lead terms for five days before the treatment. We control for product fixed effects and daily fixed effects.

Table 27: DiD Estimates of the Impact of PMG on Prices with Leads à la Autor (2003).

Products Prices (log)	(1) DiD	(2) DiD
$T_{i,l,t} * P_{i,t-1}$	-0.0121	-0.0117
	(0.0187)	(0.0187)
$T_{i,l,t} * P_{i,t-2}$	0.0030	0.0040
	(0.0181)	(0.0181)
$T_{i,l,t} * P_{i,t-3}$	0.0067	0.0076
	(0.0191)	(0.0191)
$T_{i,l,t} * P_{i,t-4}$	0.0155	0.0162
	(0.0199)	(0.0198)
$T_{i,l,t} * P_{i,t-5}$	0.0205	0.0211
	(0.0219)	(0.0218)
Observations	9,028	9,028
R-squared	0.986	0.986
Controls	NO	YES
Product Dummies	YES	YES
Time Dummies	YES	YES
F Test	0.842	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Figure 10: DiD Estimates of the Impact of PMG on Prices (Full sample, with Controls) with Leads à la Autor (2003).



ducing artificially timed treatments and artificially treated subjects. Subjects and treatments fake assignments are drawn from two Bernoulli distributions

with parameters p (probability of success) derived from the sample distributions of $Treated_{i,l,t}$ and $Post_{i,t}$ respectively. Within this setting, we should not observe any significant effect of PMG repeals on prices. Comfortingly, results reported in Table 28 confirm this prediction.
Next, we conduct another falsification test by estimating our models after substituting the dependent variable with a placebo outcome that should not be affected by PMG shutdown. In particular, we generate fake product prices drawn by random distributions resembling sample ones (same mean and variance). Results shown in Table 29 confirm the absence of any impact of PMG repeals on fake outcome.

Table 28: DiD and DDD Estimates of the Impact of *Fake* Implementation Period on Prices for *Fake* Treated/Control Samples.

Products Prices (log)	(1) DiD	(2) DiD	(3) DDD	(4) DDD
$T_{i,l,t} * P_{i,t}(Fake)$	0.0018	0.0019		
	(0.00262)	(0.00261)		
$T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}(Fake)$			0.0033	0.0033
			(0.00403)	(0.00403)
Observations	9,028	9,028	9,028	9,028
R-squared	0.986	0.986	0.986	0.986
Controls	NO	YES	NO	YES
Product Dummies	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES
F Test	0.502	0.000	0.418	0.000

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in in parentheses *** p<0.01, ** p<0.05, * p<0.1

In order to analyze if our main results are robust to the exclusion of a particular product we estimate the baseline model (15) after dropping one product at a time. Results suggest that this is not the case and confirm all previous findings.[78] In the same spirit, we estimate equation (15) after balancing the panel dataset and all results are confirmed.[79] Finally, it is worth noting that results do not change if we compute bootstrapped standard errors at product level.

6 Conclusions

In this work we empirically investigate the effects of Price Matching Guarantees (PMG) commercial policies on U.S. online consumer electronics prices by

[78]Results, not reported, are available from the authors upon request.

[79]Precisely, we drop first 34 days in which we observe only some products; results are available upon request.

Table 29: DiD and DDD Estimates of the Impact of PMG on *Fake* Prices.

Fake Products Prices (log)	(1) DiD	(2) DiD	(3) DDD	(4) DDD
$T_{i,l,t} * P_{i,t}$	-0.0011	-0.0011		
	(0.00154)	(0.00154)		
$T_{i,l,t} * P_{i,t} * HRHV_{i,l,t}$			-0.0013	-0.0013
			(0.00203)	(0.00204)
Observations	9,028	9,028	9,028	9,028
R-squared	0.999	0.999	0.999	0.999
Controls	NO	YES	NO	YES
Product Dummies	YES	YES	YES	YES
Time Dummies	YES	YES	YES	YES
F Test	0.466	0.722	0.516	0.746

Notes: All specifications include time and product fixed effects. Controls include product popularity, number of reviews and rating. Robust Standard Errors in in parentheses *** p<0.01, ** p<0.05, * p<0.1

applying a Difference-in-Difference research design.

Estimates conducted over a sample of product prices, observed on the NewEgg platform between May and October 2018, provide evidence in favor of an average price reduction of about 3.9% after the interruption of the PMG policy. In order to have a more detailed picture of the issue, we investigate if such effect is heterogeneous across products. In particular, we focus on products features that might affect the outcome of PMG policies and that can be recovered exclusively on online markets. Platform data allow us to obtain information from User Generated Contents (UGC), like product popularity, product quality and online search visibility (Google Search Rank); indeed, we believe that these product characteristics might provide indirect information on consumers heterogeneity. Estimates conducted on specific subsamples show that when PMG are interrupted, low quality (low search rank) products experience a price increase of about 3.4%, while for high quality (high visibility) products a price reduction of about 3.7% is observed.

These findings are in line with the lack of unambiguous predictions of the theoretical literature and are consistent with models predicting anti-competitive effects of PMG policies and with those interpreting such policies as a price discriminating device. Theoretical models predicting anti-competitive effects of PMG, suggest that such policies might induce higher prices in oligopoly markets (as the online consumer electronics) by sustaining collusion. In particular, online retailing platforms can easily monitor competitors prices trough price-tracking systems and can react faster to price signals, if compared to brick and mortar retailers. This possibility might sustain collusion by decreasing information asymmetries among competitors and reducing detection lags. On the other side, buyers' sensitivity to product quality and the willingness to engage in search activity can indirectly identify those customers whose demand is more rigid, thus allowing price discrimination practices. Indeed, e-commerce allows platforms to easily recover information on buyers, thanks also to UGC, thus favoring discrimination policies.

Models that predict anti-competitive effects of PMG on prices are well suited to explain the results for high quality and visible products. The demand of such products is high and stable and consumers are likely to be available to pay a price premium. Such features, together with easily detectable price signals,make collusion more sustainable. Thus, PMG policies might be an invitation to collude that can be quickly and easily captured by competitors. However, it is worth noting that our analysis does not allows us to support such theoretical interpretation of the results since we do not analyse NewEggs competitors' behavior.

Our empirical results are also consistent with theoretical models arguing that PMG act as price discrimination tools. Indeed, such theoretical explanation requires a significant percentage of consumers invoking PMG rights; unfortunately, we do not have data on PMG redemption frequency. However, Moorthy and Winter [28] find redemption rates ranging between 5% and 25% on a sample of 46 retailers operating in the United States and in Canada. It is reasonable to assume that online markets redemption rates can be similar to physical ones, thus providing support to the price discrimination interpretation of PMG policies.

References

[1] Arbatskaya, M., Hviid, M. and Shaffer, G. [2000], 'Promises to match or beat the competition: Evidence from retail tire prices', advances in applied microeconomics (advances in applied microeconomics, volume 8)'.

[2] Arbatskaya, M., Hviid, M. and Shaffer, G. [2006], 'On the use of low-price guarantees to discourage price cutting', *International Journal of Industrial Organization* **24**(6), 1139–1156.

[3] Autor, D. H. [2003], 'Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing', *Journal of labor economics* **21**(1), 1–42.

[4] Baye, M. R. and Kovenock, D. [1994], 'How to sell a pickup truck:'beat-or-pay'advertisements as facilitating devices', *International Journal of Industrial Organization* **12**(1), 21–33.

[5] Belton, T. M. [1987], 'A model of duopoly and meeting or beating competition', *International Journal of Industrial Organization* **5**(4), 399–417.

[6] Byrne, D. P. and De Roos, N. [2019], 'Learning to coordinate: A study in retail gasoline', *American Economic Review* **109**(2), 591–619.

[7] Cabral, L., Dürr, N., Schober, D. and Woll, O. [2018], Price matching guarantees and collusion: Theory and evidence from germany, Technical report, Working Paper, New York University.

[8] Cameron, A. C. and Miller, D. L. [2015], 'A practitioner's guide to cluster-robust inference', *Journal of human resources* **50**(2), 317–372.

[9] Chen, Z. [1995], 'How low is a guaranteed-lowest-price?', *Canadian Journal of Economics* pp. 683–701.

[10] Chilet, J. A. [2018], 'Gradually rebuilding a relationship: The emergence of collusion in retail pharmacies in chile'.

[11] Chung, H. S., Kim, M. et al. [2016], 'Low-price guarantees and pricing behavior: evidence from hypermarkets in korea', *Economics Bulletin* **36**(2), 1223–1229.

[12] Coad, A. [2009], 'On the distribution of product price and quality', *Journal of Evolutionary Economics* **19**(4), 589–604.

[13] Constantinou, E. and Bernhardt, D. [2018], 'The price-matching dilemma', *International Journal of Industrial Organization* **59**, 97–113.

[14] Corts, K. S. [1997], 'On the competitive effects of price-matching policies', *International Journal of Industrial Organization* **15**(3), 283–299.

[15] Doyle, C. [1988], 'Different selling strategies in bertrand oligopoly', *Economics Letters* **28**(4), 387–390.

[16] Edlin, A. S. [1997], 'Do guaranteed-low-price policies guarantee high prices, and can antitrust rise to the challenge?', *Harvard Law Review* pp. 528–575.

[17] Ellison, G. and Ellison, S. F. [2009], 'Search, obfuscation, and price elasticities on the internet', *Econometrica* **77**(2), 427–452.

[18] Gorodnichenko, Y. and Talavera, O. [2017], 'Price setting in online markets: Basic facts, international comparisons, and cross-border integration', *American Economic Review* **107**(1), 249–82.

[19] Haruvy, E. and Leszczyc, P. T. P. [2016], 'Measuring the impact of price guarantees on bidding in consumer online auctions', *Journal of Retailing* **92**(1), 96–108.

[20] Hay, G. A. [1981], 'Oligopoly shared monopoly and antitrust law', *Cornell L. Rev.* **67**, 439.

[21] Hess, J. D. and Gerstner, E. [1991], 'Price-matching policies: An empirical case', *Managerial and Decision Economics* **12**(4), 305–315.

[22] Hviid, M. and Shaffer, G. [1999], 'Hassle costs: the achilles' heel of price-matching guarantees', *Journal of Economics & Management Strategy* **8**(4), 489–521.

[23] Hviid, M. and Shaffer, G. [2010], 'Matching own prices, rivals'prices or both?', *The Journal of Industrial Economics* **58**(3), 479–506.

[24] Jain, S. and Srivastava, J. [2000], 'An experimental and theoretical analysis of price-matching refund policies', *Journal of Marketing Research* **37**(3), 351–362.

[25] Logan, J. W. and Lutter, R. W. [1989], 'Guaranteed lowest prices: do they facilitate collusion?', *Economics Letters* **31**(2), 189–192.

[26] Lu, Y. and Wright, J. [2010], 'Tacit collusion with price-matching punishments', *International Journal of Industrial Organization* **28**(3), 298–306.

[27] Mago, S. D. and Pate, J. G. [2009], 'An experimental examination of competitor-based price matching guarantees', *Journal of Economic Behavior & Organization* **70**(1-2), 342–360.

[28] Moorthy, S. and Winter, R. A. [2006], 'Price-matching guarantees', *The RAND Journal of Economics* **37**(2), 449–465.

[29] Moorthy, S. and Zhang, X. [2006], 'Price matching by vertically differentiated retailers: Theory and evidence', *Journal of Marketing Research* **43**(2), 156–167.

[30] Nalca, A., Boyaci, T. and Ray, S. [2010], 'Competitive price-matching guarantees under imperfect store availability', *Quantitative Marketing and Economics* **8**(3), 275–300.

[31] OECD [2019], *An Introduction to Online Platforms and Their Role in the Digital Transformation.*
URL: *https://www.oecd-ilibrary.org/content/publication/53e5f593-en*

[32] Png, I. P. and Hirshleifer, D. [1987], 'Price discrimination through offers to match price', *Journal of Business* pp. 365–383.

[33] Pollak, A. et al. [2017], Do price-matching guarantees with markups facilitate tacit collusion? theory and experiment, Technical report.

[34] Salop, S. C. [1986], Practices that (credibly) facilitate oligopoly coordination, *in* 'New developments in the analysis of market structure', Springer, pp. 265–294.

[35] Stallkamp, M. and Schotter, A. P. [2019], 'Platforms without borders? the international strategies of digital platform firms', *Global Strategy Journal* .

[36] Wilhelm, S. [2016], Price-matching strategies in the german gasoline retail market, Technical report, Working Paper, Goethe Universität Frankfurt, erhältlich unter: http://ssrn ....

[37] Wu, C., Wang, K. and Zhu, T. [2015], 'Can price matching defeat showrooming', *University of California, Haas School of Business, Berkeley* .

[38] Zhuo, R. [2017], 'Do low-price guarantees guarantee low prices? evidence from competition between amazon and big-box stores', *The Journal of Industrial Economics* **65**(4), 719–738.

A Appendix

Table 30: Sub-Categories List.

Sub - Categories	# products
CPU Processor	3
Computer Case	2
Mobile Phone	1
Scanner	2
Speaker	2
Motherboard	1
Monitor	3
Headset	1
USB Flash	1
CPU Cooler	1
Speaker for Domotic	1
Tablet	1
Desktop PC	1
Laptop PC	1
Power Supply	1
Printer	2
Memory Card	2
Hard Disk	1
Smart Thing Domotic	2

Table 31: Products List.

Products Titles
AMD Ryzen 5 1500X Processor
Corsair Crystal Series 570X RGB - Tempered Glass; Premium ATX Mid-Tower Case
BlackBerry PRIV (32GB) Verizon Factory Unlocked Phone
Fujitsu fi-7160 Color Duplex Document Scanner
Fujitsu ScanSnap S1300i Instant PDF Multi Sheet-Fed Scanner
Philips BT50B/37 Wireless Portable Bluetooth Speaker
Asus ROG MAXIMUS VIII FORMULA DDR4 ATX Motherboards
ASUS VS247H-P 23.6 Full HD 1920x1080 2ms HDMI DVI VGA Monitor
Samsung Hmd Odyssey Windows Mixed Reality Headset
Samsung 128GB BAR (METAL) USB 3.0 Flash Drive
Corsair CW-9060025-WW Hydro Series Liquid CPU Cooler
Echo Dot (2nd Generation) - Smart speaker with Alexa - Black
ASUS VivoMini Mini PC
Dell XF9PJ Latitude 7490 Notebook
Intel Core i7-8700 Desktop Processor 6 Cores
AMD Ryzen 7 2700X Processor Wraith Prism LED Cooler
Corsair RMx Series RM850 x 80 PLUS Gold Fully Modular ATX Power Supply
ASUS 24-inch Full HD FreeSync Gaming Monitor
Brother Monochrome Laser Printer; Compact All-in One Printer
Team 64GB microSDXC UHS-I/U1 Class 10 Memory Card with Adapter
LG Electronics 21.5 Screen LED-Lit Monitor
HP LaserJet Pro M227fdw All-in-One Wireless Laser Printer
Logitech Z313 Speaker System + Logitech Bluetooth Audio Adapter Bundle
PNY CS900 960GB 2.5 Sata III Internal Solid State Drive (SSD)
Samsung SmartThings ADT Wireless Home Security Starter Kit
Samsung SmartThings Smart Home Hub
Rosewill 2U Server Chassis Server Case (RSV-2600)
Corsair Apple Certified 16GB (2 x 8GB) DDR3 1333 MHz (PC3 10600) Laptop Memory
Acer Iconia One 10 NT.LDPAA.003 10.1-Inch Tablet