

How Does Big Data Facilitate Price Discrimination?

Yingqi Ge¹, Ruotong Huang², Yibang Zhang^{3*}

¹School of Economics, University of Sichuan, Chengdu, China

²Wuhan Britain-China School, Wuhan, China

³School of Professional Studies, University of Columbia, New York, USA

**Corresponding Author. Email: yibangzhang221@gmail.com*

Abstract. In modern society with advanced technology, big data has become an important production tool. The development of big data has led to the emergence of a large number of online sales industries while also bringing other effects, such as price discrimination. Big data can effectively empower price discrimination, which will have a significant impact on different entities. Therefore, this article will focus on the process of empowering price discrimination with big data, explore the steps or methods required for the formation of price discrimination in the era of big data, and then pay attention to the differences or advantages of discriminatory pricing under big data compared to traditional sales models.

Keywords: Big data, Price discrimination, Algorithm, Model, Pricing Decision

1. Introduction

Since the 21st century, the continuous development of science and technology in modern society has brought great convenience to people's lives with the emergence of networks and mobile devices. In the era of rapid technological development, big data has emerged, changing people's traditional way of life and creating new online sales models. It has been welcomed by people for its flexibility and convenience and has largely replaced physical stores as one of the most important symbols of the 21st century. In addition, in the past few years, data has become a crucial factor of production, and information asymmetry has become an important reason for the competitiveness of individuals and businesses. The collection and processing of data can help businesses better understand customer preferences and expectations and make corresponding judgments and decisions based on them. For example, how to recommend suitable products to consumers, whether to update their own products in a timely manner and, more importantly, whether big data can be used to squeeze consumers, that is price discrimination. This article's question unfolds here, illustrating how big data actually drives price discrimination.

The literature search studied the relationship between big data and price discrimination, but due to the relatively short development history of big data, the number of researchers and published papers is not sufficient. Currently, this article literature lacks sufficient data, and basic case analysis is used as a substitute.

The research of big data is closely related to people's lives, and any information left by people on the internet may be collected for price discrimination. Therefore, the main purpose of this article is

to clarify the process and mechanism of big data empowering price discrimination and discuss the reasons for big data empowering price discrimination. With these previous studies, this paper prescribes the right medicine, analyzes the impact of discriminatory pricing on social welfare in the context of big data, alerts people to pay attention to price discrimination in the era of big data and raises awareness of prevention, and provides corresponding research assistance for regulatory authorities.

This article provided a brief summary and induction of the literature they searched for. The literature review section of Section 2 follows the logical sequence of "Big Data Empowering Price Discrimination - How Big Data Empowers Price Discrimination - The Advantages of Price Discrimination Empowered by Big Data - Some Case Studies." The following Section 3 is a specific study of the case. This article has found a significant example in the Chinese consumer market, "Didi Taxi", which uses electronic devices used by consumers to discriminate against prices and investigates the phenomenon of price discrimination against Netflix. Section 4 is the results section, which studies the results of big data empowering price discrimination through case analysis; Section 5 is the conclusion section, which summarizes the important viewpoints of the entire text and provides theoretical explanations. Finally, the limitations section of Section 6 focuses on the shortcomings of this article, reflecting on them and expressing hope.

The main research direction of this article is "How big data promotes price discrimination", while also paying attention to some potential issues, providing valuable suggestions for the further development of human society and technological progress.

2. Literature review

The concept of price discrimination was first proposed by British economist Pigou in 1920 and has a history of over a hundred years. However, price discrimination does not seem easy to implement in traditional sales, and there are many reasons for this, such as uniform pricing or consumer deception, which will be explained in detail in the following conclusion. Nowadays, data has become an important factor of production, and the development and mining of data has become a cutting-edge topic. Data technology can lead to various discriminatory and unfair outcomes, which often increasingly harm the interests of vulnerable groups [1]. Among numerous discriminatory outcomes, price discrimination is the most influential typical representative, and the era of big data is empowering the "promotion" of price discrimination. Martínez wrote: With the continuous development of technology, large companies will use the information left by people on the Internet to judge consumers' personal behavior and preferences and finally purposefully "guide" consumers into the personalized pricing consumption trap they designed [2]. Wang uses examples to illustrate, and list multiple types of price discrimination case studies (such as Travel Apps, Ctrip, and Internet Shopping) to expose the fact that big data empowers price discrimination [3]. Ji focuses on exploring the behavioral nature and regulation of "big data-enabled price discrimination against existing customers", and elaborates on the historical process of big data promoting the development of price discrimination [4]. Price discrimination in the era of big data has a certain degree of "concealment", because big data can use information asymmetry to amplify competitive and related advantages, and consumers often do not know that their personal data is widely used [5].

Price discrimination in the era of big data is essentially no different from the traditional sense of price discrimination, but it surpasses it in terms of its generation and manifestation, which is precisely the focus of this study. FRA believes that the application of price discrimination in big data is a form of decision automation in which algorithms are used to execute command sequences on computers to obtain inputs and generate outputs [6]. Zhao revealed the fundamental reason for price

discrimination in the context of big data - the irrationality and opacity of algorithm pricing and also pointed out that the three behavioral elements that constitute big data price discrimination include establishing algorithm programs, collecting user data, and discriminatory pricing, and this is consistent with the former's viewpoint [7]. However, what the article refers to as 'big data promoting price discrimination' is a universal viewpoint, and it does not mean that there are not a few special situations that often require the participation of 'external forces'. Yu and Jia demonstrated through experiments that although the essence of price discrimination is information asymmetry between ordinary customers (RC) and e-commerce platforms (E-CP), not all big data can empower price discrimination: in e-commerce platforms, data portability and consumer sensitivity to price changes can to some extent suppress price discrimination behavior [3]. Choe et al. also proposed that merchants can obtain data through data tracking tools (such as cookies) or direct purchases, but excessive data collection can actually intensify competition and cause mutual harm [8].

Big data promotes price discrimination, not only facilitating the process of "price discrimination", but also promoting the results or outputs of discriminatory pricing. The use of big data algorithms has increased the possibility of implementing price discrimination, and the differential pricing methods of platforms comply with this implicit discrimination standard but may not necessarily violate anti-monopoly regulations while also increasing production [9]. Some papers suggest that price discrimination in big data can have welfare effects, which can promote an increase in profitability. Taking Netflix's service as an example, after achieving effective customized pricing, sellers will be able to approach consumers' booking tools and participate in what we call almost perfect price discrimination [10]. Divide the concept of price discrimination into three categories based on degree, among which the most common and easy to implement are second-degree price discrimination and third-degree price discrimination. However, some scholars have proposed that with the advent of the era of big data, first-degree price discrimination is easier to realize today [11]. The "almost perfect price discrimination" mentioned earlier refers to the theoretical first-degree price discrimination. From this perspective, big data has almost perfectly promoted price discrimination.

Finally, this article also studied cases from different countries: Wang demonstrated how price discrimination brought about by big data affects the online transactions that modern people excel at through his research on many different online trading platforms and social media in China [12]; Woodcock analyzed an example of a US video website to illustrate how companies analyze big data and use price discrimination to increase profits, as well as the consequences of big data price discrimination [13].

3. Case analysis

Against the backdrop of rapid technological development in modern society, the phenomenon of price discrimination is becoming increasingly common, especially in today's era where online transactions are prevalent. The operation of this situation cannot be separated from the support of big data. So, what is big data, what is the connection between big data and price discrimination, and how does big data drive price discrimination? This article will provide specific explanations through the following articles and examples.

Many people believe that big data is just a term representing the personal information left behind by people while browsing various web pages or using electronic products. But in fact, Martínez article provides a more detailed description of big data, which is actually a general term for a series of behaviors related to information, including the collection and use of user personal information, how to analyze this information to make it better utilized, and how to classify and summarize the

collected information [2]. This series of behaviors are all descriptions of big data as a whole. So, how does big data promote price discrimination? Zhao showed us the connection between big data and price discrimination [14]. Big data can help businesses collect consumer information faster and more accurately, including but not limited to various online consumption websites, social media platforms, and other channels. Enterprises organize and analyze the collected data, drawing curves like statues of real users, in order to make targeted judgments on each consumer and implement personalized pricing, also known as price discrimination. Martínez suggests that in addition to extensive personalized pricing, businesses can also use big data information, such as consumers' personal preferences, to develop highly targeted advertising [2].

The following example will provide a detailed explanation of how big data discriminates against consumers in terms of price in daily life. In Wang's article, there are two very common examples of price discrimination based on big data [3]. The first example is a consumer who discovers that the price difference between purchasing a plane ticket from the Ctrip App and the official HNA App exceeds 1500 RMB, and there are also abnormal fluctuations in ticket prices when purchasing a plane ticket from the Ctrip App. Ctrip App is a travel app commonly used by Chinese people. Many people will buy air tickets, train tickets, or book hotels from it, which gives enterprises a lot of personal preference data to support their analysis and decision-making so as to make price discrimination against individuals. This is the most common example of big data driving price discrimination, which collects information through commonly used online consumption software and utilizes people's low sensitivity to online consumption to implement price discrimination.

The second example of Wang is that companies will discriminate against consumers based on their region [3]. As mentioned in the article, Ms. Wang's online shopping products will have different prices when delivered to different regions. The price difference between delivering products to CBD areas and industrial buildings is about 30%, which is a very surprising figure. This kind of price discrimination is even more obvious. In addition to using personal consumption preferences, big data also formulates more detailed personalized pricing based on the consumption level of the region where the consumer is located. From this perspective, companies have specially marked the areas where wealthy people are located, which is equivalent to designing a special model. When big data recognizes this information, price discrimination will occur. A similar example is Wang ride-hailing app, which sets different prices based on consumers' mobile phone systems [12]. In addition to consumer consumption records, big data can also promote price discrimination in many ways.

Moreover, this article will take the Netflix platform as an example for case analysis. Today, search engines like Netflix, Amazon, and Yelp have introduced personalization to their websites, giving users personalized suggestions no matter the movie, product, or business review. Apparently, personalization improves users' experience by increasing search efficiency and the quality of content, especially when the user searches an ambiguous query like "apple". Based on the user's previous history of searches and purchases, Amazon is able to return more relevant products like electronic devices rather than fruits, which increases the likelihood that the user will buy them [15]. However, by utilizing users' personal information, these e-commerce websites have an economic incentive to ask wealthy users or users with higher demand to pay higher prices for the same products. According to Branco and Brossard-Ruffey, information posted while browsing on the Internet exposes a user's geographic location, financial situation, application, and browser history to the content of communications, the variety and quantity of which constitute a digital portrait of each person [10]. Sellers can use it to accurately estimate a customer's retention utility or at least their degree of preference for competing goods.

Netflix, as a large video browser, exemplifies this concept. Although Netflix uses secondary price discrimination, which sets different prices according to the number of purchases made by consumers, the article uses Ordering -Choice Model Averaging (OMA) to estimate Netflix's profit from personalized pricing [16]. Big data makes it possible to estimate the value of an individual Netflix subscription by capturing nearly 5,000 web viewing variables at low prices.

To start with, Netflix has a favorable environment to achieve price discrimination. First, because Netflix users' purchases and subscriptions are made online, Netflix can offer customized prices more efficiently than in physical stores. Secondly, it can be seen from Netflix's use of second-degree price discrimination that Netflix has the ability to collect customer information for price discrimination. Third, Netflix can do empirical research on hypothetical personalized pricing, which is not possible in many other situations. This requires individual-level data, including web browsing history and purchases of all specific items, which are hard to come by at the same time. However, Netflix subscriptions can be easily inferred through users' browsing history.

Shiller used the Probit model to analyze the role of web browsing data in predicting which consumers are likely to subscribe to Netflix [17]. Without any additional information, the probability of each person subscribing is about 16%. However, when the model included standard demographic data such as race, age, income, and whether or not they had children, the predictive power increased, and the range of subscription probabilities for individuals widened to 6 to 30 percent. Finally, the OMA approach introduced a complete set of variables, including web browsing history, to the model, which significantly improved the prediction accuracy, with subscription probabilities ranging from close to 0% to 99.8%.

Then, using an empirical model to correlate the improved predictive accuracy from web browsing data to key outcome variables, pricing, and profit can be imitated under the counterfactual situation that occurs when Netflix implements first-degree price discrimination. The study found that web browsing behavior significantly increased profit growth from first-degree price discrimination compared to second-degree price discrimination - a 12.18% increase in profits when all data was used for customized pricing and only 0.79% when demographic data was relied on alone. As a result, web browsing data makes personalized pricing more attractive to companies and more likely to be implemented, influencing consumer behavior. Overall, consumer surplus is expected to shrink by 8%, raising serious equity concerns - some consumers may pay twice as much as others for the same product.

The above example briefly describes how big data drives price discrimination. Simply put, businesses collect people's daily life information, integrate this information, and design it into exquisite algorithm models so that they can make discriminatory pricing decisions with input and output. Of course, big data can do far more than that.

4. Results

The above case analysis provides direction for this research. In the context of big data, people's personal information is compiled into valuable data, and differentiated pricing is based on data, which is known as price discrimination in the era of big data.

Firstly, for data demanders, the data they need should reflect the identity characteristics and utility value of consumers. Therefore, they often capture information that consumers unintentionally or intentionally leave behind while browsing products, which aggregates together to form a digital portrait of the user, allowing merchants to more accurately evaluate consumers' expectations or retention value. After collecting a certain amount of data, it will be fed into a specific model, which is the algorithm design. Through algorithms, different weights or parameters can be assigned to data

information based on their importance, forming a "black box" that only needs to wait for the next user to leave data. The "black box" can then operate on its own and generate a specific result, which is the price that meets the user's utility. Of course, algorithm design cannot be fabricated out of thin air but requires data training to fit real-world situations as closely as possible before it can be widely used.

Secondly, for the implementation environment of price discrimination, online sales are undoubtedly a better choice. Online sales can only bring data which is incomparable to physical stores. The smooth development of online sales also relies on the support of technology and digital technology, so this article can easily draw the important conclusion that big data can promote price discrimination from here. Of course, this point will also be elaborated in detail in the conclusion of later chapters.

Price discrimination can increase the profits of online platforms. Different platforms adopt different types of price discrimination based on their own platform and product characteristics. This is the fundamental reason why online platforms are adopting big data for price discrimination, and it is also an important reason for merchants or algorithm designers to constantly innovate their programs. A small browsing history can also leave a significant impact, which is a masterpiece of the big data era. Although this may sound a bit vulgar, the fact is that it is.

5. Limitation

Firstly, there are still shortcomings in the research on how big data can promote price discrimination in this article. Due to the fast development speed and short time of big data, there are relatively few references available. Therefore, the author can only analyze and summarize the relationship between big data and price discrimination through limited literature and concepts. In some places, there are problems, such as inadequate explanation and insufficient depth of elaboration. Secondly, this article is written based on concepts and existing cases, without collecting a large amount of data on specific consumers for research and analysis and without using professional software for programming and design, lacking strong evidence. Finally, this study is only limited to the process of big data promoting price discrimination and does not elaborate too much on the consequences of big data causing price discrimination, namely the impact on people's welfare and the policy responses for supervision, management, and prevention. These are the difficulties that need to be overcome in the future and the key issues that people need to pay attention to in the era of big data.

6. Conclusion

This research topic is how big data can promote price discrimination, so the research background is set in the era of big data. The main features of the big data era include but are not limited to explosive growth in data volume, skyrocketing data processing speed, deep mining of data value, and severe challenges to data security. Therefore, consumption in the era of big data is no longer the same as before, and users will leave traces whether browsing or purchasing products - significantly different from traditional consumption's "incognito shopping". So, this article can first draw the following conclusion: price discrimination is not easy to establish in traditional consumption, but in today's world, where data has become a factor of production, price discrimination can easily turn from theory to reality. In the traditional sales industry dominated by offline sales, there are several reasons why price discrimination is difficult to implement. Firstly, the traditional retail industry lacks strong data statistical capabilities, making it difficult to record, identify, store, and utilize consumer consumption information. If the information master in the era of big data is the merchant,

then the information master in the traditional sales model is the consumer. Anonymous consumption makes it difficult for merchants to capture useful information, especially for large supermarkets, which presents greater challenges for data integration. Secondly, consumers can deceive businesses by hiding personal information to prevent price discrimination. For example, consumers can wear inexpensive clothes to small vendors to purchase goods - compared to wearing expensive clothes, they will largely not face price discrimination, or tourists who pretend to be locals will largely not be "exploited." Thirdly, traditional retail industry has clear pricing, and once the price is indicated on the product, it cannot be changed in a short period of time. Even for some products that are not clearly priced, such as vegetables and meat in the market, consumers can bargain and reduce potential discriminatory pricing to normal or even lower prices. Finally, due to the competitive environment and regulatory restrictions, similar products often adopt uniform pricing, such as restaurants specializing in scenic spots where the prices of all similar dishes are exactly the same. Otherwise, they will be "betrayed" by consumers or isolated by other businesses. But in the era of big data, many of the above restrictions and barriers have been broken, and data has given new magic to sales models, making price discrimination seem "easy".

Going back to now, data is invisible or intangible, and simply collecting data does not lead to price discrimination. So, how can big data promote the implementation of price discrimination? From scattered data to the phenomenon of price discrimination, several steps need to be taken. Step 1: Data collection and organization. In the era of big data, consumption has shifted from offline to online, making it easier to obtain data such as consumer purchase records, browsing habits, and personal information. Merchants can install data tracking tools (such as cookies) or purchase data from data brokers and then summarize and integrate the data. One thing to note here is that collecting more data does not necessarily mean better. On the contrary, data collection should be focused and selective, with a tendency to collect data on consumer purchase history and preference choices in order to make the next round of pricing decisions. Excessive data may lead to "over-identification", while insufficient data may lead to "omission bias". In addition, there may be certain biases in the data itself, such as selecting only a portion of population data and ignoring other factors, coupled with data quality issues, which may lead to data bias. Incorrect data cannot accurately generate user profiles. Step 2: Algorithm design and processing. With the help of the product of the intelligent era - computers, algorithm design is carried out, and designers form specific "formulas" by inputting explanatory variables, dependent variables, and collected data. Among them, the explanatory variable and the dependent variable may not have a causal relationship, but more importantly, they need to have a correlation, which enables the algorithm to generate certain results for pricing decisions after inputting new data. Like data collection work, the programming quality, technology, language, etc. of algorithms can affect the output results of algorithms, which may also lead to the emergence of some "biases". Step 3: model training and adjustment. After the algorithm design is completed, it cannot be immediately put into production use and still needs to be trained. The purpose of training is to make the model as realistic as possible, reduce potential problems and risks, and lower the supervision time and cost for algorithm designers. After inputting a large amount of data, evaluate the output results to observe whether they meet consumers' purchasing intentions and utility and whether they can generate more profits. It is worth noting that the data also has priority, and the weight of important data is greater, which will greatly affect the output results. Therefore, attention should also be paid to the issue of "bias". If errors are found during the training process, certain adjustments and modifications need to be made to optimize the model and algorithm itself. With the rapid development of the times, the speed of model optimization is also accelerating, and the significance of model training is becoming more

important. Step 4: Decision making. According to the output of the algorithm, merchants will bundle and classify consumers or product quantities and provide different pricing services for different entities, which is known as second-degree price discrimination and third-degree price discrimination. After a complete process, price discrimination appears around people. The price discrimination brought about by big data may lead to a series of problems, such as personalized pricing, fair competition, changes in consumer decision-making, and market transparency, which in turn invisibly affect price discrimination and consumer behavior choices.

According to theoretical knowledge, price discrimination can be divided into three categories based on their degree: first-degree price discrimination, second-degree price discrimination, and third-degree price discrimination. But in reality, second-degree price discrimination and third-degree price discrimination are the most common because they do not have high requirements for data volume, data processing capabilities, etc. In the era of big data, information and data are crucial production factors that empower productivity, promote continuous iteration and upgrading of products and industries, and impact people's consumption concepts. With the development of science and technology, data analysis and processing capabilities, as well as algorithm technology, will continue to innovate and upgrade. In the intelligent and digital future society, fully personalized first-degree price discrimination is expected to be achieved, and individual humans will become massive amounts of data constantly being mined in the scientific world.

Acknowledgement

Yingqi Ge, Ruotong Huang, and Yibang Zhang contributed equally to this work and should be considered co-first authors.

References

- [1] M. Favaretto, E. De Clercq and B. S. Elger, "Big Data and Discrimination: Perils, Promises and Solutions. A Systematic Review, " *Journal of Big Data*, vol. 6, no. 1, 2019. doi: 10.1186/s40537-019-0177-4
- [2] S. Martínez, "Price discrimination and big data: Highest prices every time we click?" (Master's thesis, Law and Economics Master's Programme, 2022).
- [3] Z. Wang, "Classic Case Study of Price Discrimination in the Age of Big Data, " *Highlights in Business, Economics and Management*, vol. 20, pp. 42 - 47, 2023. doi: 10.54097/hbem.v20i.12317
- [4] L. Ji, "Legal regulation of price discrimination based on big data in the context of digital economy, " *Open Journal of Legal Science*, vol. 12, no. 6, pp. 4129 - 4136, 2024. doi: 10.12677/ojls.2024.126586
- [5] M. Gambaro, "Big Data Competition and Market Power, " *Market and Competition Law Review*, vol. 2, no. 2, pp. 99 - 122, 2019. doi: 10.7559/mclawreview.2018.326
- [6] European Union Agency for Fundamental Rights, "Big data, algorithms and discrimination, " *Publications Office of the European Union*, 2018. doi: 10.2811/315832
- [7] Q. Zhao, "Exploration on the protection of consumer rights and interests in big data price discrimination, " *International Journal of Frontiers in Sociology*, vol. 6, no. 4, pp. 15 - 20, 2024. doi: 10.25236/IJFS.2024.060403
- [8] C. Choe, S. King and N. Matsushima, "Pricing with cookies: Behavior-based price discrimination and spatial competition, " *Management Science*, vol. 64, no. 12, pp. 5669 - 5687, 2018, doi: 10.1287/mnsc.2017.2873.
- [9] X. Zhang, "Study on Antitrust Regulation of Price Discrimination on Big Data Platforms, " *Academic Journal of Business & Management*, vol. 5, no. 27, 2023. doi: 10.25236/ajbm.2023.052715
- [10] S. Branco and A. Brossard-Ruffey, "WELFARE EFFECTS OF PRICE DISCRIMINATION: THE IMPACT OF BIG DATA, " 2017. <https://repositori.upf.edu/bitstream/handle/10230/33468/BrancoBrossard%20TFM2017.pdf?isAllowed=y&sequence=1>
- [11] Y. Wang, "Application and Analysis of Price Discrimination in China's E-commerce based on Big Data Analysis, " in *2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*, pp. 1690 - 1693, Atlantis Press, 2022.

- [12] Y. Wang, "Price Discrimination in the Era of Big Data, " Wwww.atlantis-Press.com; Atlantis Press, 2022. doi: 10.2991/assehr.k.220504.102
- [13] R. A. Woodcock, "Big data, price discrimination, and antitrust, " Hastings LJ, vol. 68, p. 1371, 2016..
- [14] X. Zhao, "Big data and price discrimination, " in 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), pp. 471 - 475, IEEE, 2020.
- [15] A. Hannak, G. Soeller, D. Lazer, A. Mislove and C. Wilson, "Measuring Price Discrimination and Steering on E-commerce Web Sites, " Proceedings of the 2014 Conference on Internet Measurement Conference - IMC '14, 2014. doi: 10.1145/2663716.2663744
- [16] J. Yu and W. Jia, "Research on price discrimination behavior governance of e-commerce platforms: A Bayesian game model based on the right to data portability, " Axioms, vol. 12, no. 10, p. 919, 2023. doi: 10.3390/axioms12100919
- [17] B. Shiller, "First-Degree Price Discrimination Using Big Data, " 2014. https://www.brandeis.edu/economics/RePEc/brd/doc/Brandeis_WP58R2.pdf