IMPACT OF DATA POLICY ON PLATFORM ECONOMY

YU LINGFEI

MPhil

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Logistics and Maritime Studies

Impact of Data Policy on Platform Economy

YU Lingfei

A thesis submitted in partial fulfilment of the

requirements for the degree of Master of Philosophy

May, 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_________________________ (Signed)

YU Lingfei (Name of student)

Abstract

The proliferation of data-driven personalization in digital platforms has become increasingly critical in today's platform economy. This focus on leveraging user data plays a pivotal role in enhancing user engagement and optimizing platform service quality, while navigating the complex landscape of user privacy concerns. As users' awareness of data privacy increases, regulatory policies represented by GDPR require individuals to have greater control over their personal data, which affects the platform's data collection strategies. However, implementing effective data policies is not straightforward. Platforms face distinct challenges and trade-offs when balancing personalization benefits with privacy protections. In practice, we observe three primary data collection scenarios: (1) Data-Free Collection ($\mathbb{N}$), where platforms rely on generalized recommendations without user data; (2) Mandatory Collection ($\mathbb{M}$), where platforms collect comprehensive user data for personalized recommendations; and (3) Voluntary Collection ($\mathbb{V}$), where users opt-in to data sharing, balancing personalization with privacy. In this thesis, we aim to investigate the impacts of different data collection scenarios on a monopolistic platform's recommendation strategies, pricing decisions, and user demand. We analyze the platform's decisions and performance under each data collection scenario and evaluate their implications for user surplus and platform profitability, focusing on two main chapters: the impact of data policies on user polarization and platform service level.

Our analysis yields several key insights. (1) First, the effectiveness of recommendation strategies and user polarization tolerance significantly influence the platform's optimal data collection choice. In scenarios with high privacy sensitivity, Scenario $\mathbb{N}$ may outperform by leveraging generalized recommendations, while Scenario $\mathbb{M}$ excels in contexts where personalization drives engagement. Scenario $\mathbb{V}$ often balances these trade-offs, maximizing consumer surplus by aligning personalization with user consent. (2) Second, Scenario $\mathbb{V}$ frequently achieves a win-win outcome by enhancing service quality while respecting privacy preferences. (3) Third, higher recommendation accuracy under Scenario $\mathbb{M}$ may not always benefit users if privacy concerns increase searching costs, whereas Scenario $\mathbb{N}$ can mitigate this in privacy-sensitive markets by reducing the searching cost. Our results provide actionable managerial insights for platform operators, emphasizing the need to carefully select data collection strategies that align with user preferences and market conditions to optimize engagement and profitability.

Acknowledgements

Over the years, PolyU has steadfastly embodied its motto, "*To learn and to apply, for the benefit of mankind.*" Rooted in Hong Kong with a global perspective, PolyU has consistently put this motto into practice, fostering a promising future for education and innovation. As I near the completion of my studies, I wish to express my heartfelt gratitude to my supervisors, friends, and everyone who has supported me along this journey.

I am deeply grateful to my chief supervisor, Prof. Guo Xiaomeng, whose exceptional academic expertise has earned her widespread respect. Prof. Guo's meticulous care, respect, and dedication to her students are admired by all students in LMS. I sincerely wish for Prof. Guo to continue advancing global scientific research and achieving even greater milestones. My life experiences shines in these years as Prof. Guo's kind education.

I am equally thankful to Prof. Xiao Guang for his unwavering support and encouragement. His optimistic outlook on life has been a constant source of inspiration, and his spirited determination on the sport has motivated me to face life's challenges with courage and positivity.

I owe Prof. Wang Xiaoyu too much. Words cannot fully express my appreciation for his mentorship. I aspire to contribute meaningfully to society after graduation, as this would be the greatest way to honor Prof. Wang's influence. As one of the most impactful emerging scholars I have encountered, Prof. Wang's genuine care and dedication have earned him the admiration of all his students.

Additionally, I am immensely grateful to PolyU for providing a dynamic platform for learning and personal growth. Professors such as Prof. SU Nancy, Prof. Eric and Dr. Patrick have opened numerous opportunities for me to explore, Ms Priscilla also helps me a lot. In my personal life, Ms. Zhang Xiaoli, Dr. Zhao Xue, Mr. Xu Zheng, Ms. Wang Yajing, Mr. Zhuang Lei, Ms. Zhu Wenqian, Ms. Tan Hongjun, and Ms. Kryshell have provided invaluable support, enriching my experiences and making my journey at PolyU truly fulfilling.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Digital Privacy

User data is an invaluable asset in the realm of platform recommendations, as it enables platforms to create detailed user profiles that enhance the personalization of content. By analyzing user interactions, preferences, and behaviors, platforms can tailor recommendations to align closely with individual interests, thereby increasing user engagement and satisfaction. This data-driven approach not only helps in delivering a more personalized user experience but also allows platforms to choose strategies that maximize their profits.

However, the extensive use of user data also presents significant risks, particularly concerning data misuse and abuse. The potential for data leakage is a major concern (Fainmesser, Galeotti, and Momot 2023b), as it can lead to unauthorized access to sensitive user information, resulting in privacy breaches and identity theft. Data breach examples include Aadhaar in India[1], First American

[1]BBC: https://www.bbc.com/news/world-asia-india-42575443

Financials[2] in USA and Cambridge Analytica scandal in Facebook (CNN 2018[3]). Moreover, the misuse of user data for purposes beyond the user's consent, such as unauthorized sharing with third parties, can erode trust and damage the platform's reputation. These risks underscore the importance of implementing robust data security protocols and fostering transparency in data handling practices.

The General Data Protection Regulation (GDPR[4]) has been instrumental in shifting data rights back to users, empowering them to have greater control over their personal information. Under GDPR, users have the right to opt-in or opt-out of data collection practices, ensuring that their consent is obtained before their data is processed. This regulation mandates that platforms provide clear and accessible information about how user data is collected, used, and shared, thereby enhancing transparency and accountability. For instance, GDPR requires platforms to implement data protection by design and default, conduct data protection impact assessments, and appoint data protection officers to oversee compliance. These measures aim to safeguard user privacy and ensure that data handling practices align with users' rights and expectations.

Users must trade-off between data profiling and data control rights, as platform recommendations heavily rely on user data. Specifically: (i) While data profiling enables platforms to deliver personalized content (i.e., personalized recommendation strategy) and enhance user experience, it also necessitates the sharing of personal information, which can compromise privacy and potentially lose of receiving new contents. (ii) Conversely, exercising data control rights, such as opting out of data collection, may forcing the platform to provide recommendations

[2]https://www.cenfedcu.org/landing/first-american-financial-data-breach
[3]CNN: https://edition.cnn.com/2018/05/02/politics/cambridge-analytica-closure/index.html
[4]GDPR: https://gdpr-info.eu/

that meets the major users (i.e., generalized recommendation strategy), potentially diminishing user satisfaction as platform cannot know the user preference. This interaction between data profiling and data control is crucial, as it influences the balance between personalization and privacy. Users must weigh the benefits of personalized recommendations against the potential risks to their privacy, making informed decisions about their data sharing preferences.

1.2 User Polarization

Social media platforms like TikTok have transformed content consumption by utilizing sophisticated recommendation systems that rely heavily on user data. These platforms collect vast amounts of information from users to personalized content recommendations that align with individual preferences. This data-driven approach is essential for maintaining user engagement, as it allows platforms to deliver personalized content experiences that resonate with users' unique interests and behaviors.

The user data collected by platforms such as TikTok is instrumental in providing precisely personalized content recommendations, a process often referred to as polarization (Santos, Lelkes, and Levin 2021). Polarization offers significant benefits: (i) for users, it means encountering content that closely matches their interests, reducing their searching efforts, more focus on contents they prefer; (ii) for the platform, it results in increased user retention and engagement, as users are more likely to remain active when they consistently find appealing content. This can lead to longer attention durations and more frequent interactions, fostering a vibrant user community. Additionally, precise recommendations enable platforms to target advertisements more effectively, boosting advertising revenue. (Practice in polarization technology) In recent years, numerous algorithms have been developed to enhance recommendation accuracy, a field that has matured significantly in both technology and business practice (as exemplified by the Netflix Prize competition[5], 2022 ACM Dressipi RecSys Challenge[6] and Amazon KDD Cup 2023[7]). Algorithms represented by content-based filtering and collaborative filtering algorithms both infer user preferences from information about user characteristics and previous browsing behavior (J. Liu and Cong 2023). Business cases involves "People you may know" or "Whom to follow" suggestions in facebook, instagram and twitter (Shardanand and Maes 1995). These methods confine individuals to self-confirming feedback loops, resulting in intellectual isolation (J. Liu and Cong 2023), echo chambers (Cinelli et al. 2021) and popularity bias in recommendations.

While precise recommendations offer significant benefits, they also present notable drawbacks. From the platform's perspective, the negative aspects include the risk of creating echo chambers, where users are exposed only to content that reinforces their existing beliefs, potentially limiting their exposure to diverse viewpoints. This can lead to a homogenized user experience, reducing the platform's appeal to a broader audience. From the user's perspective, the negative side includes the potential for intellectual isolation, where users are confined to a narrow range of content, hindering their ability to discover new interests or perspectives. This lack of diversity in content can diminish user satisfaction over time and may

[5]Competition theme: promoting Netflix's recommending algorithm accuracy

[6]Challenge goal: to make recommendations that respond to what the user is doing during the current session. https://recsys.acm.org/recsys22/challenge/

[7]Challenge topic: Session-based recommendation, which utilizes customer session data to predict their next purchase. https://kdd.org/kdd2023/

even lead to decreased engagement.

To prevent highly similar items from being clustered together (polarized), diversity metrics for recommendation algorithms are equally important. Representative algorithm examples include ranking-based techniques in MovieLens (Ashkan et al. 2015), Determinantal Point Process (DPP) algorithm in Youtube and Hulu (Wilhelm et al. 2018; L. Chen, G. Zhang, and Zhou 2018), For You feed in TikTok[8]. Alternative business cases are both Apple Music's "Discovery Station" function, which only recommends music that the user has never heard before based on their tastes, and the recent debate over the "Start Menu Recommended Section" in Windows 11, which primarily recommends products that are irrelevant to the user's preferences. These initiatives aim to broaden users' content exposure, helping to mitigate the effects of echo chambers and intellectual isolation. By incorporating diverse recommendations, these platforms strive to enhance user satisfaction and engagement by offering a more varied and enriching content experience.

However, achieving the right balance between precise and diverse recommendations is a complex challenge for platforms. On one hand, precise recommendations are crucial for maintaining user engagement and satisfaction by delivering content that aligns closely with individual preferences. On the other hand, diverse recommendations are essential for fostering a more balanced and enriching user experience, preventing the negative effects of echo chambers and intellectual isolation.

[8]TikTok: https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

1.3 Platform Service

In the digital economy, the quality of platform services has emerged as a pivotal determinant of consumer choice and satisfaction. Superior user interface design, seamless operational experiences, and precise content recommendations significantly enhance consumer surplus—the disparity between perceived value and actual attention fee. For instance, platforms like TikTok and YouTube leverage intuitive interaction frameworks, intelligent content segmentation, and real-time feedback mechanisms to minimize users' information search costs. TikTok's "immersive swipe" interface captivates user attention by prioritizing content engagement, while YouTube's algorithmic recommendation system efficiently aligns user preferences with its vast video repository through historical viewing data analysis. Such user-centric service enhancements not only amplify consumer satisfaction but also streamline decision-making processes, enabling users to maximize utility within constrained timeframes.

The primary objective of elevating service quality is to foster a virtuous cycle of user retention and engagement. As central nodes in multi-sided markets, digital platforms must continuously refine their services to sustain user loyalty, thereby facilitating monetization and ecosystem growth. Short video platforms, for example, boost social interaction through features such as 'likes' and 'comments', which stimulate user-generated content and participation. Similarly, e-commerce platforms deploy intelligent chatbots and real-time logistics tracking to reduce service response times, thereby reducing searching cost and finally reflected by serves quality improvement.

However, the interaction between service quality and data privacy introduces a complex dynamic. The availability of user data directly shapes the precision and cost-efficiency of platform services. Data-driven personalization, such as Netflix's customized movie recommendations or Spotify's curated playlists, is based on extensive behavioral data, allowing platforms to deliver enhanced services at marginal costs. This tension creates a dynamic cost-benefit trade-off: while covert data acquisition techniques (e.g., cookie tracking or cross-platform data sharing) enable platforms to predict user needs with precision, they may diminish incentives to invest in fundamental service improvements.

A deeper paradox arises when privacy regulations constrain the quality of the service. The General Data Protection Regulation (GDPR) of the European Union, for example, requires explicit user consent for data collection, which can reduce the granularity of available data and alter the accuracy of the recommendation. This trade-off is particularly pronounced in sensitive sectors where platforms face challenges in delivering high-value, personalized services without comprehensive user data. For example, a health management platform lacking detailed biometric data may need to allocate substantial resources to developing generic health solutions, compromising service personalization. Consequently, platforms must navigate a delicate balance between data utilization and service innovation, offering critical insights into the privacy-service quality dilemma in the digital economy.

Chapter 2

Literature Review

The paper contributes to two active interdisciplinary streams of the literature: (i) digital privacy and (ii) platform recommendation

2.1 Data privacy

Interest in the data privacy literature has flourished as the users' activity data exponentially explosions in recent years. The existing literature can be separated into two distinct streams.

(1) Our model is closely related to the literature on consumer's endogenous privacy choice. Montes, Sand-Zantman, and Valletti 2019 shows the resulting inefficiencies of exclusivity data deals when users endogenous their privacy by paying a privacy cost. Johnson, Shriver, and Goldberg 2023 consider welfare implications of an opt-out model in an advertise targeting context. Choi, Jerath, and Sarvary 2023 study targeted advertising problem with consumer privacy choices. Closely related to our work, Ichihashi 2020 show that the seller prefers to commit to not

use information for pricing in order to encourage user's information disclosure.

(2) The second strand of related literature deals with the impact of privacy regulations on platforms' operations. Johnson, Shriver, and Goldberg 2023 illustrates how market structure evolves post-GDPR. Closely related to our work, F. Xu, Xiaoyu Wang, and F. Zhang 2025 investigated how the GDPR policy's transfer of data control rights would affect the supplier, retailer, and consumer welfare.

2.2 Platform Recommendation

Our research is related to the emerging literature on platform recommendation.

(1) Our model of polarization builds on the literature on (i) fairness in recommendation. Technically, researchers have explored various approaches to enhance fairness, such as incorporating fairness constraints into the recommendation algorithms (Rastegarpanah, Gummadi, and Crovella 2019) or using post-processing techniques to adjust recommendations (Pitoura, Stefanidis, and Koutrika 2022). These methods aim to balance the trade-off between accuracy and fairness, ensuring that all users receive equitable treatment. In the context of polarization, fairness in recommendation systems can play a pivotal role in either exacerbating or alleviating divisions.(Y. Wang et al. 2023; Zehlike, Yang, and Stoyanovich 2022). Our model of (ii) echo chambers refers to an environment where individuals are exposed predominantly to opinions and information that reinforce their existing beliefs, while alternative perspectives are minimized or excluded (Ge et al. 2020). This phenomenon is often facilitated by algorithms in social media and recommendation systems that prioritize content based on user preferences and past interactions. Cookson, Engelberg, and Mullins 2023 highlights the effect of echo

chambers to intensify polarization by creating insular communities where dissenting views are rarely encountered. This can lead to a reinforcement of biases and a deepening of ideological divides, as individuals become more entrenched in their viewpoints. Donkers and Ziegler 2021 shown that echo chambers can contribute to the spread of misinformation, as false or misleading information is less likely to be challenged within these closed networks.

(2) Our model of profiling builds on the literature on (i) precised recommendation, for the purpose of personalized pricing (Valletti and Wu 2020; X. Li, Xin Wang, and Nault 2024), price discriminate (Koh, Raghunathan, and Nault 2017; Ke and Sudhir 2023), queueing (D. Liu, Sarkar, and Sriskandarajah 2010) and (ii) biased recommendation (L. Li, J. Chen, and Raghunathan 2018; Qian and Jain 2024; Donnelly, Kanodia, and Morozov 2024). Closely related to our work, Bergemann and Bonatti 2024 attempted to study the interaction between heterogeneous consumers and multi-product sellers under the framework of privacy governance. They found that the platform efficiency can be improved based on the information exploitation.

2.3 Platform Service

The literature on platform service quality has gained prominence as digital platforms increasingly serve as central hubs for user interactions in the platform economy. This stream of research examines how service quality (Huang, Lyu, and Y. Xu 2022) drives consumer surplus, user retention(Mai, Hu, and Pekeč 2023), and platform profitability (Jin Li et al. 2023). Our study contributes to this literature by analyzing how data collection policies shape platform service strategies, particularly in balancing personalization with privacy constraints.

A substantial body of work highlights the pivotal role of service quality in fostering user engagement and loyalty. Smedlund 2012 introduces the concept of value co-creation, where platforms enhance service quality by facilitating interactions between users and content providers, thereby strengthening network effects. For instance, short-video platforms like TikTok leverage intuitive interfaces and real-time feedback mechanisms to minimize user search costs and boost engagement (Zhong 2023). Similarly, Huang, Lyu, and Y. Xu 2022 explore quality regulation on two-sided platforms, demonstrating that superior service quality, such as seamless matching between users and providers, enhances consumer surplus and platform profitability. These studies underscore the importance of dynamic service capabilities, including responsive design and content curation, in sustaining user satisfaction in multi-sided markets.

Recent research further investigates how service quality influences user behavior through personalization. Jun Li and Netessine 2020 analyze matching rates on online platforms, showing that high-quality matching algorithms increase user retention by aligning content with preferences. Mai, Hu, and Pekeč 2023 extend this by examining user conduct management, where platforms optimize service quality by moderating user interactions to enhance engagement. These findings align with our thesis's focus on how data-driven personalization, enabled by different data collection scenarios, affects service quality and user demand. However, these studies often assume unrestricted data access, overlooking the impact of privacy constraints on service delivery, a gap our research addresses.

The interaction between service quality and data privacy introduces significant complexity, as platforms rely on user data to deliver personalized services.

D. Liu, Sarkar, and Sriskandarajah 2010 investigate resource allocation for content personalization, demonstrating that data availability directly enhances service efficiency by reducing user search costs. However, Fainmesser, Galeotti, and Momot 2023a highlight the risks of data misuse, such as breaches or unauthorized sharing, which can erode user trust and diminish perceived service quality. Our study builds on this literature by systematically comparing service quality outcomes across three data collection scenarios: Data-Free Collection ($\mathbb{N}$), Mandatory Collection ($\mathbb{M}$), and Voluntary ($\mathbb{V}$). Unlike prior work, which often focuses on single data scenarios or unrestricted data environments, we investigate how platforms adapt service strategies to varying levels of data availability and user privacy preferences.

Despite the rich literature on platform service, several gaps remain. First, few studies explicitly examine how data collection policies influence service quality strategies in a monopolistic platform setting. While Huang, Lyu, and Y. Xu 2022 and Jun Li and Netessine 2020 address service quality in competitive or two-sided markets, they do not explore the monopoly context central to our model, where data policy directly shapes recommendation and pricing decisions. Second, the interplay between service quality and user polarization is underexplored. Although Dinerstein et al. 2018 considers search cost dynamics, the role of polarization tolerance in shaping service outcomes remains largely unaddressed. Third, existing research often adopts static frameworks, neglecting the dynamic evolution of user privacy preferences and their impact on service strategies over time.

Our thesis addresses these gaps by analyzing how data policies affect platform service quality in a dynamic, two-stage game-theoretic model. We contribute by (i) evaluating service quality across Data-Free, Mandatory, and Voluntary collection scenarios, (ii) incorporating user polarization tolerance as a key determinant of service effectiveness, and (iii) examining the dynamic interplay between data availability, privacy constraints, and service outcomes. By doing so, we provide actionable insights for platform managers to optimize service strategies in privacy-sensitive markets, enhancing both user engagement and profitability.

Chapter 3

Impact of Data Policy on User Polarization

3.1 Model Framework

3.1.1 Firms

We consider a data collection type (*Data policy*, Table 3.1) that specifies between three schemes: (i) Data free collection ($\mathbb{N}$); (ii) Mandatory collection ($\mathbb{M}$); (iii) Voluntary collection ($\mathbb{V}$). The key difference between the three schemes is that the data sets the platform collects from users.

Table 3.1: Recommendation cases in different data policy

Type	**Data-free Collection**	**Mandatory Collection**	**Voluntary Collection**
Polarize	Government website	Teladoc Health (telemedicine APP)	"For You Feed" in TikTok
Diversity	"Start Menu Recommended Section" in Windows 11	"reCAPTCHA" in Google	"Discovery Station" in Apple Music

A monopoly platform (he) decides his *recommendation strategy* based on the user's history data. We categorize his recommendation into two types according to the users' history data: (i) Personalized recommendation (users' historical data available) — the platform recommend products that will result in the highest demand to users based on the available history data, i.e., either recommend i or j; (ii) Generalized recommendation (users' historical data not available) — The platform will recommend a single product i or j to any user who choose not to disclose data, depending on which product brings him the highest demand.

3.1.2 Users

The user (she) decides whether to join the services provided by the platform (click or not, *Privacy decision*). Without loss of generality, we assume that consumers are homogeneous for the product value and uniformly distributed over the interval [0,1].

The user type is distinguished by two dimensions, as illustrated in Figure. 3.1: (1) *Preference type* — users have different preferences for the two product types *cat* (i) and *dog* (j) the platform offers. On the one hand, we define v as the base value a user would obtain when the recommendation matches his preference. On the other hand, a disutility θs would incur if the recommendation mismatches her preference, $\theta \in [0, 1]$ represents a coefficient for mismatch penalty $s, s > v$. Specifically, we define the user is a ***broad-interested user*** when $\theta \to 0$, whereas the user is a ***dedicated user*** when $\theta \to 1$. Moreover, to model the preference heterogeneity in the market, we assume that there exists γ proportion users have preference type i, while the remaining fraction $(1 - \gamma)$ proportion have preference type j. (2) *Data history* — Each user has different watching history, either i or

j. Users with fraction $\alpha, \alpha \in [0, 1]$ have watching history i, while the remaining fraction $(1 - \alpha)$ have history j.

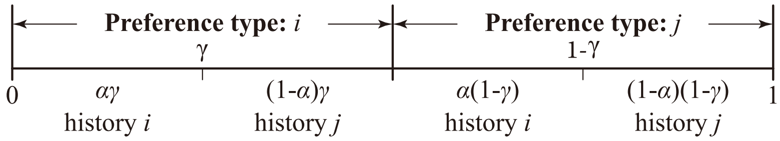


Figure 3.1: User categories.

A key feature of our model is that users have different polarization tolerance. We consider a continuous polarization tolerance level $\delta, \delta \in [0, 1]$, which is a coefficient for the base value v. The polarization tolerance level measures the users' aversion to being recommended to the same product. Specifically, we define the user is *polarization-averse* when $\delta \to 0$, whereas the user is *polarization-seeking* when $\delta \to 1$.

Moreover, we capture users' *privacy decision* in the model setup. Users with heterogeneous preference types and historical data would incur different utilities when recommended to product i or product j, which is related to the platform's personalized or generalized recommendation decision. In what follows, we use subscript $\langle H, P \rangle$ to indicate the user type, where $H, H = \{i, j\}$ represents the user's watching history and $P, P = \{i, j\}$ represents the user's preference type. Superscript $k, k = \{i, j\}$ indicates the platform's recommendation type. Given the utility categories, we can characterize the user utility $u^k_{\langle H,P \rangle}$ with different privacy decision as follows (as shown in Table. 3.2):

3.1.3 Sequence of Events

There are two stages in our model, with detailed game sequences given below.

Table 3.2: Utility $u^k_{\langle H,P\rangle}$ categories

The user's preference type (P)	Product i		Product j	
Data history (H)	i	j	i	j
Recommendation type $k=i$	$u^i_{\langle i,i\rangle}=\delta v$	$u^i_{\langle j,i\rangle}=v$	$u^i_{\langle i,j\rangle}=\delta v-\theta s$	$u^i_{\langle j,j\rangle}=v-\theta s$
Recommendation type $k=j$	$u^j_{\langle i,i\rangle}=v-\theta s$	$u^j_{\langle j,i\rangle}=\delta v-\theta s$	$u^j_{\langle i,j\rangle}=v$	$u^j_{\langle j,j\rangle}=\delta v$

Stage 0 — Data policy: The nature first gives its data collection scenario (data-free collection, mandatory collection or voluntary collection).

Stage 1 — Recommendation Decision: The platform decides his product recommendations portfolio for different types of users based on the data availability.

Stage 2 — Privacy Decision: Given the recommendation algorithm observed: (i) for the data free collection scenario, the user data is unknown; (ii) for the mandatory collection scenario, users data is totally collected; (iii) for the voluntary collection scenario, the user should decide whether to disclose her data.

Stage 3 — Activity Decision (demand): Users decide whether to click after observing the platform's recommendation.

3.2 Data-free Collection Scenario ($\mathbb{N}$)

In this section, we analyze the implications of a platform economy operating without data collection, a scenario increasingly relevant in light of growing privacy concerns and regulatory scrutiny. Platforms such as DuckDuckGo, which prioritize user privacy by limiting or anonymizing data collection, serve as pertinent case studies. These platforms contrast with data-intensive counterparts like Google or Facebook, offering a unique lens to examine how data policies shape economic outcomes in the platform economy.

In the absence of data collection, platforms forego the ability to leverage user data for personalized services, targeted advertising, or third-party data sharing, which are central to the revenue models of many platform economies. For instance, DuckDuckGo, a search engine that does not track user queries, relies on contextual advertising rather than personalized data-driven ads, potentially limiting its revenue potential but enhancing user trust. According to backward induction, we complete the proposition in two steps:

3.2.1 Step 1: Users' privacy disclosing decision and activity decision

(i) *Activity Decision (demand):* given the different user utility, if and only if $u^k_{\langle H,P\rangle} > 0$ would the user join the digit service the platform offers, it derives the user activity level $d^k_{\langle H,P\rangle}$ based on utility maximize conditions (Table. 3.3).

(ii) *Privacy Decision:* In this case, since there is no historical data, given the platforms' generalized recommendation decision, the total users get the generalized surplus of either S^i or S^j.

Table 3.3: Users' surplus and demand under generalized recommendation scheme i and j in scenario N

		Preference type i users		Preference type j users	
	History	i	j	i	j
Recommend type i	Surplus $s^i_{\langle H,P\rangle}$	$\int_0^1 u^i_{\langle i,i\rangle}\,d\theta$	$\int_0^1 u^i_{\langle j,i\rangle}\,d\theta$	$\int_0^{\theta^i_{\langle i,j\rangle}} u^i_{\langle i,j\rangle}\,d\theta$	$\int_0^{\theta^i_{\langle j,j\rangle}} u^i_{\langle j,j\rangle}\,d\theta$
	Demand $d^i_{\langle H,P\rangle}$	1	1	$\bar{\theta}^i_{\langle i,j\rangle}$	$\bar{\theta}^i_{\langle j,j\rangle}$
Recommend type j	Surplus $s^j_{\langle H,P\rangle}$	$\int_0^{\bar{\theta}^j_{\langle i,i\rangle}} u^j_{\langle i,i\rangle}\,d\theta$	$\int_0^{\bar{\theta}^j_{\langle j,i\rangle}} u^j_{\langle j,i\rangle}\,d\theta$	$\int_0^1 u^j_{\langle i,j\rangle}\,d\theta$	$\int_0^1 u^j_{\langle j,j\rangle}\,d\theta$
	Demand $d^j_{\langle H,P\rangle}$	$\bar{\theta}^j_{\langle i,i\rangle}$	$\bar{\theta}^j_{\langle j,i\rangle}$	1	1

*Note that $\bar{\theta}^i_{\langle i,j\rangle} = (\delta v)/s$, $\bar{\theta}^i_{\langle j,j\rangle} = v/s$, and $\bar{\theta}^j_{\langle i,i\rangle} = v/s$, $\bar{\theta}^j_{\langle j,i\rangle} = (\delta v)/s$

The total user surplus in Data-free Collection Scenario is given by:

$$S_N^i = \underbrace{\alpha\gamma s^i_{\langle i,i\rangle} + (1-\alpha)\gamma s^i_{\langle j,i\rangle}}_{\text{Surplus of preference type } i \text{ users}} + \underbrace{\alpha(1-\gamma)s^i_{\langle i,j\rangle} + (1-\alpha)(1-\gamma)s^i_{\langle j,j\rangle}}_{\text{Surplus of preference type } j \text{ users}} \tag{3.1}$$

$$S_N^j = \underbrace{\alpha\gamma s^j_{\langle i,i\rangle} + (1-\alpha)\gamma s^j_{\langle j,i\rangle}}_{\text{Surplus of preference type } i \text{ users}} + \underbrace{\alpha(1-\gamma)s^j_{\langle i,j\rangle} + (1-\alpha)(1-\gamma)s^j_{\langle j,j\rangle}}_{\text{Surplus of preference type } j \text{ users}} \tag{3.2}$$

each of which is depended on the platform's recommendation decision.

3.2.2 Step 2: The platform's equilibrium results (recommendation decision)

The objective of the platform is a function that is maximizing users' activity (demand). In other words, the profit maximizing goals of the platform is separate from the user utility. The platform's expected profit based on generalized recommendation scheme i and j can be written as:

$$\pi_N^i = \underbrace{\alpha\gamma d^i_{\langle i,i\rangle} + (1-\alpha)\gamma d^i_{\langle j,i\rangle}}_{\text{Profits from preference type } i \text{ users}} + \underbrace{\alpha(1-\gamma)d^i_{\langle i,j\rangle} + (1-\alpha)(1-\gamma)d^i_{\langle j,j\rangle}}_{\text{Profits from preference type } j \text{ users}} \tag{3.3}$$

$$\pi_N^j = \underbrace{\alpha\gamma d^j_{\langle i,i\rangle} + (1-\alpha)\gamma d^j_{\langle j,i\rangle}}_{\text{Profits from preference type } i \text{ users}} + \underbrace{\alpha(1-\gamma)d^j_{\langle i,j\rangle} + (1-\alpha)(1-\gamma)d^j_{\langle j,j\rangle}}_{\text{Profits from preference type } j \text{ users}} \tag{3.4}$$

3.2.3 Results

To sum up, users' total surplus and the platform's profit are given by:

(i). The users' total surplus can be written as: $S_N = \begin{cases} S_N^i, & \text{if } 0 < \alpha < \bar{\alpha}_{N,1} \\ S_N^j, & \text{if } \bar{\alpha}_{N,1} < \alpha < 1 \end{cases}$

(ii). The platform's profit can be written as: $\pi_N = \begin{cases} \pi_N^i, & \text{if } 0 < \alpha < \bar{\alpha}_{N,2} \\ \pi_N^j, & \text{if } \bar{\alpha}_{N,2} < \alpha < 1 \end{cases}$

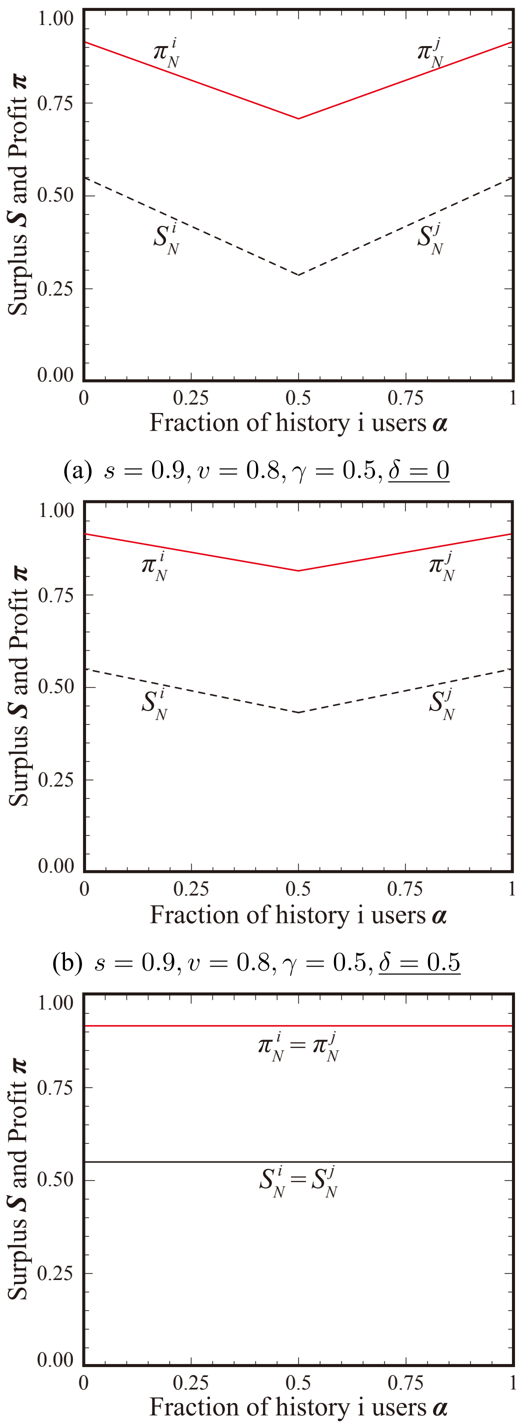
Note that $\bar{\alpha}_{N,1} = \frac{v\left(\gamma\delta^2+\gamma-1\right)-2s((\gamma-1)\delta+\gamma)}{(\delta-1)(2s+\delta v+v)}$, $\bar{\alpha}_{N,2} = \gamma + \frac{(2\gamma-1)(v-s)}{(\delta-1)v}$. In addition, for checking purposes, let $\gamma = 0.5$, it gives $\bar{\alpha}_{N,1} = \bar{\alpha}_{N,2} = 0.5$

It can be inferred that there exist a contradiction region $\min\{\bar{\alpha}_{N,1}, \bar{\alpha}_{N,2}\} < \alpha < \max\{\bar{\alpha}_{N,1}, \bar{\alpha}_{N,2}\}$ where the platform's recommend strategy contradict to the user's preference. It can be seen graphically in Figure 3.2-3.6

PROPOSITION 3.2.1. The following statements hold.

1. For the platform's profit, one can obtain that
   (a) π_N^i and π_N^j increase in δ.
   (b) π_N^i decrease in α, whereas π_N^j increase in α.
   (c) π_N^i increase in γ, whereas π_N^j decrease in γ.
2. For the users' surplus, one can obtain that
   (a) S_N^i and S_N^j increase in δ.
   (b) S_N^i decrease in α, whereas S_N^j increase in α.
   (c) S_N^i increase in α, whereas S_N^j decrease in γ.

Proposition 3.2.1 elucidates the dynamics of platform profit (π_N^i, π_N^j) and user surplus (S_N^i, S_N^j) in the Data-Free Collection Scenario ($\mathbb{N}$), where platforms lack access to user historical data and rely on generalized recommendation strategies. The proposition's findings—specifically, the positive relationship of both profit and surplus with polarization tolerance (δ), the inverse effects of historical data

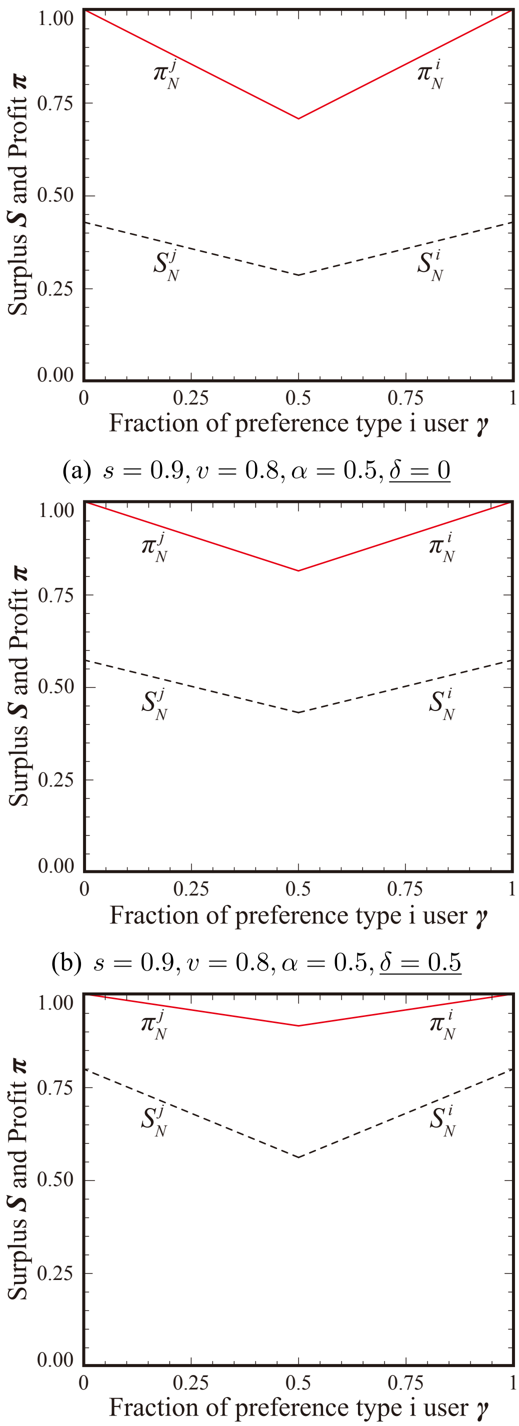


(a) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\delta = 0}$

(b) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\delta = 0.5}$

(c) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\delta = 1}$

Figure 3.2: Surplus and profit maximize segments in Data-free Collection Scenario (α versus π and S).

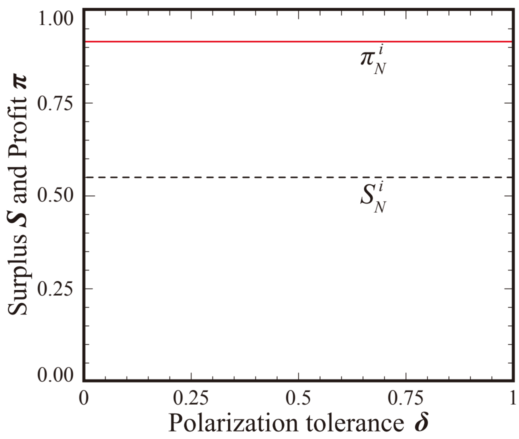


(a) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\delta = 0}$

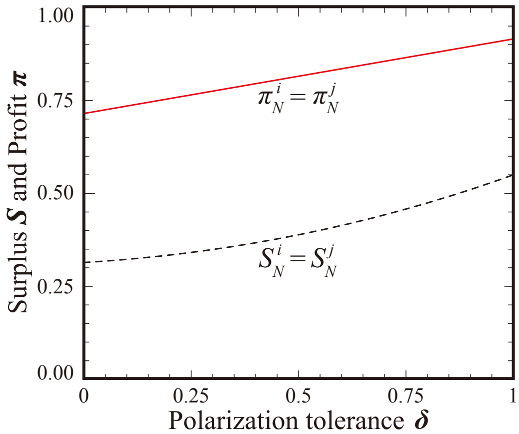
(b) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\delta = 0.5}$

(c) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\delta = 1}$

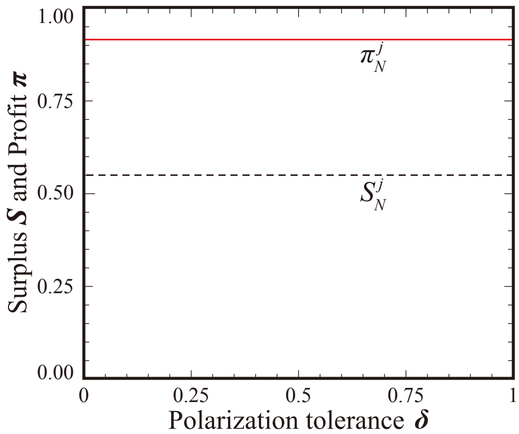
Figure 3.3: Surplus and profit maximize segments in Data-free Collection Scenario (γ versus π and S).



(a) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\alpha = 0}$

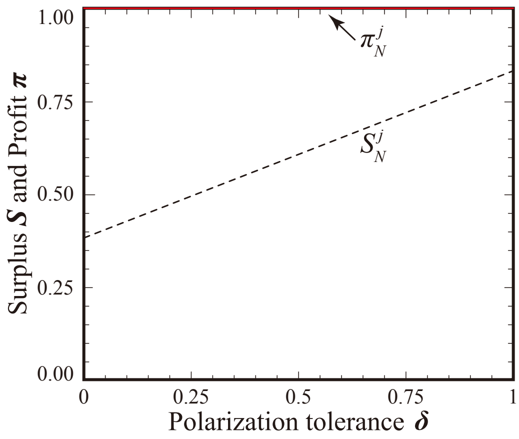


(b) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\alpha = 0.5}$

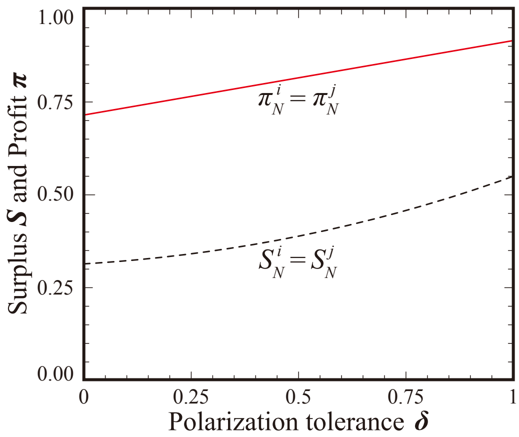


(c) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\alpha = 1}$

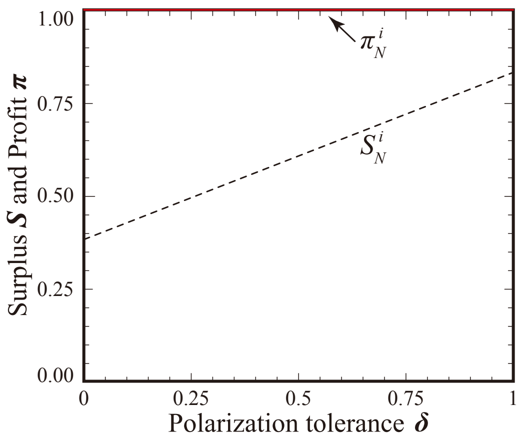
Figure 3.4: Surplus and profit maximize segments in Data-free Collection Scenario (δ versus π and S).



(a) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\gamma = 0}$

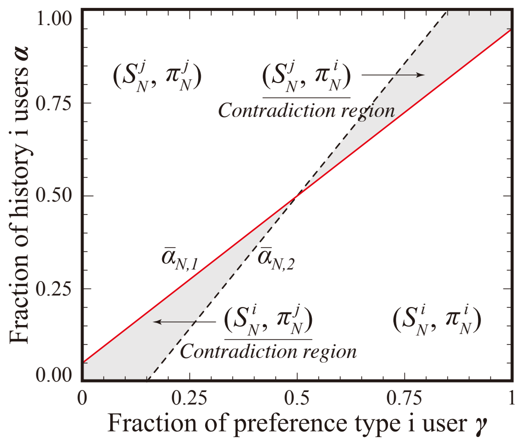


(b) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\gamma = 0.5}$

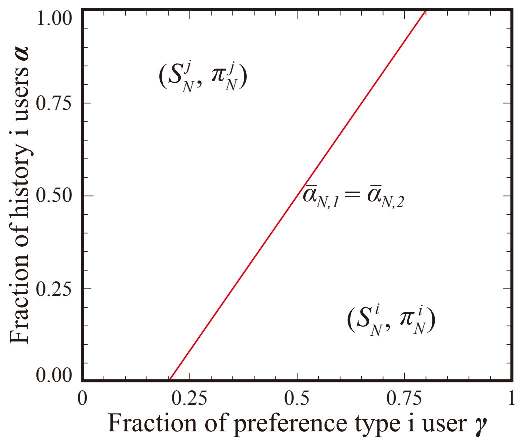


(c) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\gamma = 1}$

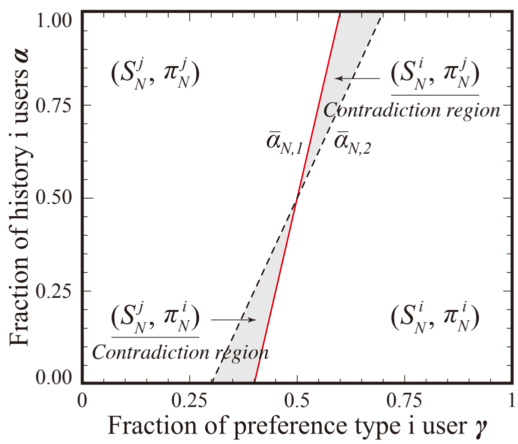
Figure 3.5: Surplus and profit maximize segments in Data-free Collection Scenario (δ versus π and S).



(a) $s = 0.9, v = 0.8, \delta = 0.4$



(b) $s = 0.9, v = 0.8, \delta = 0.66$



(c) $s = 0.9, v = 0.8, \delta = 0.9$

Figure 3.6: Surplus and profit maximize segments in Data-free Collection Scenario (γ versus α).

distribution (α) on different recommendation strategies, and the influence of preference heterogeneity (γ)—offer critical insights for platform managers navigating data-constrained environments. These insights translate into actionable strategies for optimizing service quality, user engagement, and profitability in multi-sided digital markets.

1. **Leveraging Polarization Tolerance to Enhance Engagement**: The proposition indicates that both platform profit and user surplus increase with δ, the coefficient representing users' tolerance for polarized recommendations. This suggests that platforms operating without user data should prioritize content strategies that align with users' openness to repetitive or similar recommendations. For instance, managers of short-video platforms (e.g., TikTok) can design content feeds that emphasize familiar themes or genres, as polarization-tolerant users are more likely to engage with such content, thereby boosting session durations and ad revenue. To operationalize this, platforms could employ clustering algorithms to identify user segments with high δ values based on observable proxies (e.g., repeated interactions with similar content) and tailor generalized recommendations accordingly.

2. **Adapting to Historical Data Distribution**: The inverse relationship between α (the proportion of users with history i) and π_N^i (profit from recommending product i) versus the positive relationship with π_N^j (profit from recommending product j) underscores the importance of aligning recommendation strategies with the dominant user history. Managers must assess the market's historical data distribution, even in data-free scenarios, by analyzing publicly available trends or proxy metrics (e.g., social media hashtags

or search engine trends). For example, a platform like YouTube, lacking granular user data, could infer dominant content preferences (e.g., music versus tutorials) from aggregate viewership patterns and adjust its generalized recommendation to favor the product (i or j) that maximizes demand. This strategic alignment mitigates the risk of mismatched recommendations, enhancing user satisfaction and platform profitability.

3. **Balancing Preference Heterogeneity**: The proposition highlights that π_N^i increases with γ (the proportion of users preferring product i), while π_N^j decreases, with parallel effects on user surplus. This implies that platforms must carefully adjust their recommendation strategies to reflect the market's preference heterogeneity. In markets with high γ, managers should prioritize recommending product i to capture greater demand and surplus, whereas in markets with low γ, product j becomes more viable. For e-commerce platforms like Amazon, this could involve promoting product categories (e.g., electronics versus fashion) based on regional or demographic preference trends, inferred from sales data or market research. By dynamically adjusting recommendations to match preference distributions, managers can optimize consumer surplus and reinforce network effects, increasing market demand.

4. **Navigating Profit-Surplus Mismatch**: The existence of a contradiction region (max$\{\bar{\alpha}_{N,1}, \bar{\alpha}_{N,2}\} < \alpha <$ min$\{\bar{\alpha}_{N,1}, \bar{\alpha}_{N,2}\}$) where platform recommendation strategies conflict with user preferences highlights a potential mismatch between profit maximization and user satisfaction. Managers must proactively address this by integrating user feedback mechanisms (e.g.,

surveys or engagement analytics) to detect when generalized recommendations deviate from user expectations. For instance, a platform observing declining engagement metrics could experiment with hybrid recommendation strategies that blend polarized and diversified content, mitigating the risk of user churn while maintaining profitability. This approach is particularly relevant for platforms in competitive markets, where user loyalty is critical for sustaining network effects.

5. **Strategic Implications for Data-Free Environments**: In the absence of user data, Proposition 3.2.1 underscores the need for platforms to invest in alternative data sources and inference techniques to approximate α, γ, and δ. Managers could leverage third-party market research, anonymized behavioral analytics, or partnerships with data aggregators to inform recommendation strategies. For example, a health management platform unable to collect biometric data could use population-level health trends to recommend generalized wellness programs, balancing accessibility with relevance. Additionally, investing in user interface enhancements (e.g., intuitive navigation or real-time feedback) can compensate for the lack of personalized recommendations, ensuring that service quality remains competitive even under data constraints.

3.3 Mandatory Collection Scenario ($\mathbb{M}$)

In this section, we examine the scenario of mandatory data collection within the platform economy. Mandatory data collection, where platforms require users to disclose personal data as a condition of service, is exemplified by platforms like

Facebook (now Meta), XiaoHongShu which rely heavily on user data to drive personalized services and targeted advertising. To rigorously analyze the economic and strategic implications of this policy, we employ backward induction and complete the proposition in two steps:

3.3.1 Step 1: Users' privacy disclosing decision and activity decision

(i) *Activity Decision:* given the different user utility, if and only if $u^k_{\langle H,P\rangle} > 0$ would the user join the digit service the platform offers, it derives the user activity level $d^k_{\langle H,P\rangle}$ and user surplus $s^k_{\langle H,P\rangle}$ based on utility maximize conditions.

(ii) *Privacy Decision:* In this case, the platform holds historical data (H) for all users, but cannot infer user preferences (P). Henceforth, the platform makes his personalized recommendation decision by maximizing the expected demand $\mathbb{E}[d^k_{\langle H,P\rangle}]$ for each users with different history, either $k = i$ or $k = j$.

The total user surplus in Mandatory Collection Scenario is given by:

$$S^k_M = \underbrace{\max\left\{S^i_{\langle i,P\rangle}, S^j_{\langle i,P\rangle}\right\}}_{\text{Surplus of history } i \text{ users}} + \underbrace{\max\left\{S^i_{\langle j,P\rangle}, S^j_{\langle j,P\rangle}\right\}}_{\text{Surplus of history } j \text{ users}} \tag{3.5}$$

Note that:

- $S^i_{\langle i,P\rangle} = \alpha\gamma s^i_{\langle i,i\rangle} + \alpha(1-\gamma)s^i_{\langle i,j\rangle}$;
- $S^j_{\langle i,P\rangle} = \alpha\gamma s^j_{\langle i,i\rangle} + \alpha(1-\gamma)s^j_{\langle i,j\rangle}$;
- $S^i_{\langle j,P\rangle} = (1-\alpha)\gamma s^i_{\langle j,i\rangle} + (1-\alpha)(1-\gamma)s^i_{\langle j,j\rangle}$;
- $S^j_{\langle j,P\rangle} = (1-\alpha)\gamma s^j_{\langle j,i\rangle} + (1-\alpha)(1-\gamma)s^j_{\langle j,j\rangle}$

3.3.2 Step 2: The platform's equilibrium results (recommendation decision)

The objective of the platform is a function that is maximizing users' activity $d^k_{\langle H,P\rangle}$ (user traffic). In other words, the profit maximizing goals of the platform is separate from the user utility. The platform's expected profit based on personalized recommendation can be written as:

$$\pi^k_M = \underbrace{\max\left\{D^i_{\langle i,P\rangle}, D^j_{\langle i,P\rangle}\right\}}_{\text{Profits from history } i \text{ users}} + \underbrace{\max\left\{D^i_{\langle j,P\rangle}, D^j_{\langle j,P\rangle}\right\}}_{\text{Profits from history } j \text{ users}} \tag{3.6}$$

Note that:

- $D^i_{\langle i,P\rangle} = \alpha\gamma d^i_{\langle i,i\rangle} + \alpha(1-\gamma)d^i_{\langle i,j\rangle}$;
- $D^j_{\langle i,P\rangle} = \alpha\gamma d^j_{\langle i,i\rangle} + \alpha(1-\gamma)d^j_{\langle i,j\rangle}$;
- $D^i_{\langle j,P\rangle} = (1-\alpha)\gamma d^i_{\langle j,i\rangle} + (1-\alpha)(1-\gamma)d^i_{\langle j,j\rangle}$;
- $D^j_{\langle j,P\rangle} = (1-\alpha)\gamma d^j_{\langle j,i\rangle} + (1-\alpha)(1-\gamma)d^j_{\langle j,j\rangle}$

3.3.3 Results

To sum up, as illustrated in Figure 3.7-3.10, users' total surplus and the platform's profit are given by:

(i). The users' total surplus can be written as:

$$\hat{S}_M = \begin{cases} S^j_{\langle i,P\rangle} + S^j_{\langle j,P\rangle}, & \text{if } 0 < \gamma < \bar{\gamma}_{M,1} \\ S^j_{\langle i,P\rangle} + S^i_{\langle j,P\rangle}, & \text{if } \bar{\gamma}_{M,1} < \gamma < (1-\bar{\gamma}_{M,1}) \\ S^i_{\langle i,P\rangle} + S^i_{\langle j,P\rangle}, & \text{if } (1-\bar{\gamma}_{M,1}) < \gamma < 1 \end{cases}$$

(ii). The platform's profit can be written as:

$$\hat{\pi}_M = \begin{cases} D^j_{\langle i,P\rangle} + D^j_{\langle j,P\rangle}, & \text{if } 0 < \gamma < \bar{\gamma}_{M,2} \\ D^j_{\langle i,P\rangle} + D^i_{\langle j,P\rangle}, & \text{if } \bar{\gamma}_{M,2} < \gamma < (1-\bar{\gamma}_{M,2}) \\ D^i_{\langle i,P\rangle} + D^i_{\langle j,P\rangle}, & \text{if } (1-\bar{\gamma}_{M,2}) < \gamma < 1 \end{cases}$$

Note that $\bar{\gamma}_{M,1} = \frac{2\delta s - v}{2\delta s + 2s - v\delta^2 - v}$, $\bar{\gamma}_{M,2} = \frac{s-v}{2s-\delta v - v}$

PROPOSITION 3.2.2. The following statements hold.

1. For the platform's profit, one can obtain that

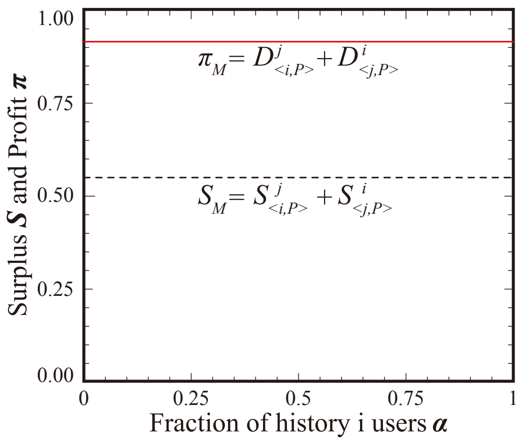
   (a) π^i_N and π^j_N increase in δ, where $\pi^k_N = D^j_{\langle i,P\rangle} + D^i_{\langle j,P\rangle}$ stay constant with δ.

   (b) π^i_N decrease in α, whereas π^j_N increase in α; and $\pi^k_N = D^j_{\langle i,P\rangle} + D^i_{\langle j,P\rangle}$ increase in $0 < \gamma < 1/2$ and decrease otherwise.

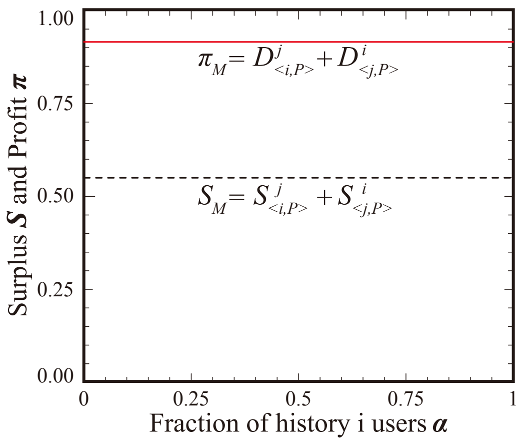
   (c) π^i_N increase in γ, whereas π^j_N decrease in γ; and $\pi^k_N = D^j_{\langle i,P\rangle} + D^i_{\langle j,P\rangle}$ increase in $0 < \alpha < 1/2$ and decrease otherwise.

2. For the users' surplus, one can obtain that

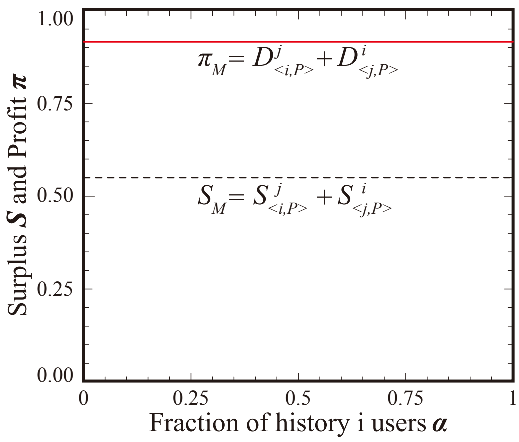
   (a) S^i_N and S^j_N increase in δ, where $S^k_N = S^j_{\langle i,P\rangle} + S^i_{\langle j,P\rangle}$ stay constant with δ.



(a) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\delta = 0}$

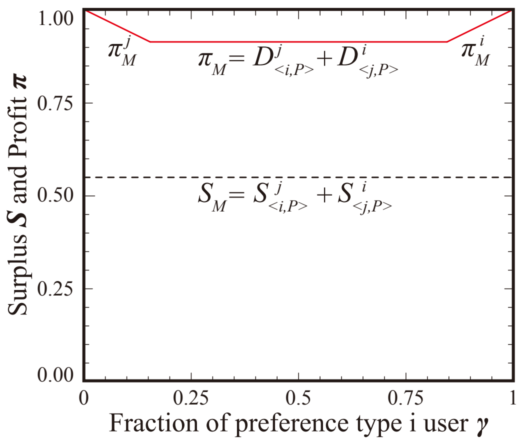


(b) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\delta = 0.5}$

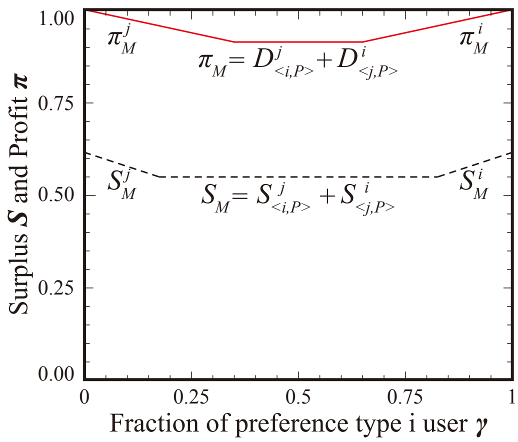


(c) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\delta = 1}$

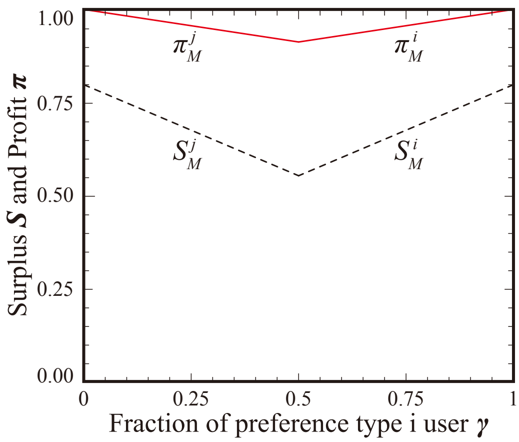
Figure 3.7: Surplus and profit maximize segments in Mandatory Collection Scenario (α versus π and S).



(a) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\delta = 0}$

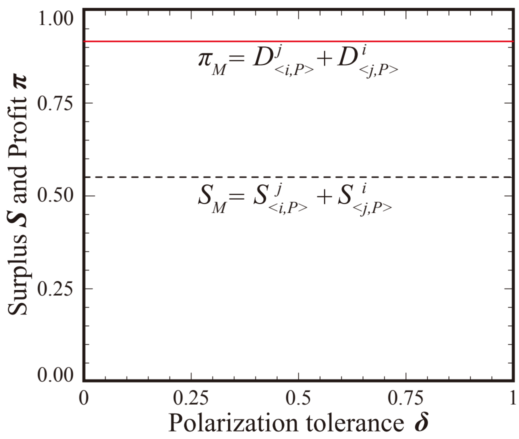


(b) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\delta = 0.5}$

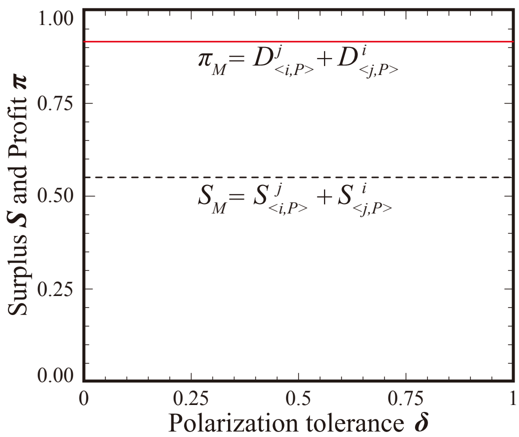


(c) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\delta = 1}$

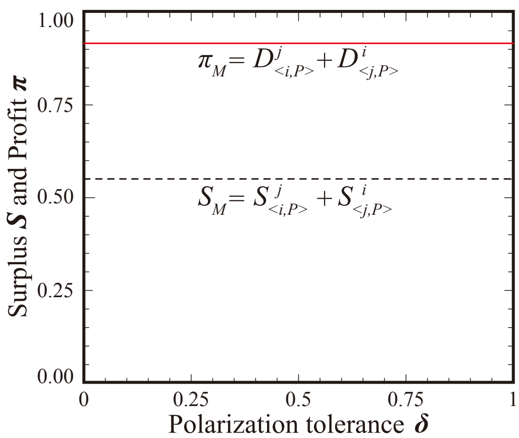
Figure 3.8: Surplus and profit maximize segments in Mandatory Collection Scenario (γ versus π and S).



(a) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\alpha = 0}$

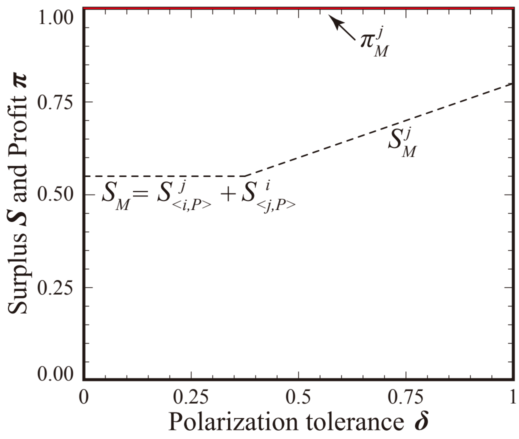


(b) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\alpha = 0.5}$

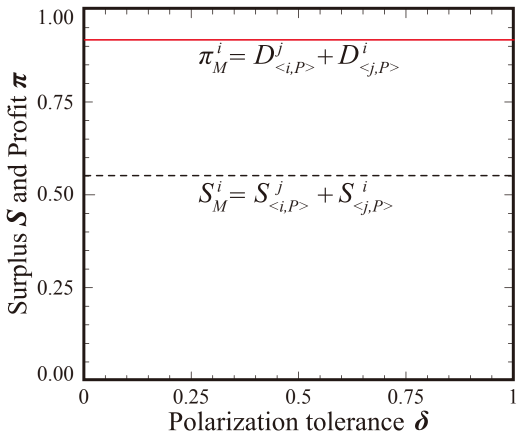


(c) $s = 0.9, v = 0.8, \gamma = 0.5, \underline{\alpha = 1}$

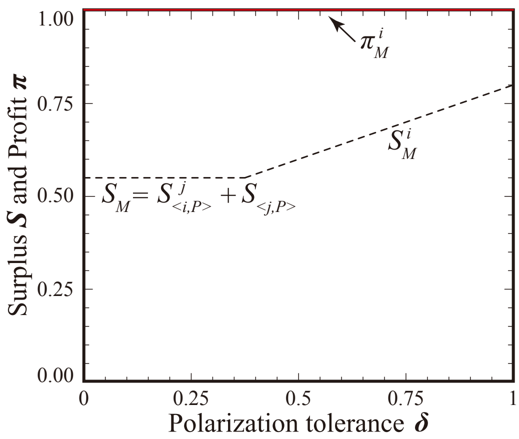
Figure 3.9: Surplus and profit maximize segments in Mandatory Collection Scenario (δ versus π and S).



(a) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\gamma = 0}$



(b) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\gamma = 0.5}$



(c) $s = 0.9, v = 0.8, \alpha = 0.5, \underline{\gamma = 1}$

Figure 3.10: Surplus and profit maximize segments in Mandatory Collection Scenario (δ versus π and S).

(b) S_N^i decrease in α, whereas S_N^j increase in α; and $S_N^k = S_{\langle i,P\rangle}^j + S_{\langle j,P\rangle}^i$ increase in $0 < \gamma < 1/2$ and decrease otherwise.

(c) S_N^i increase in γ, whereas S_N^j decrease in γ; and $S_N^k = S_{\langle i,P\rangle}^j + S_{\langle j,P\rangle}^i$ increase in $0 < \alpha < 1/2$ and decrease otherwise.

3.4 Voluntary Collection Scenario ($\mathbb{V}$)

Users have the legal rights to freely control whether their historical data is disclosed or not. Platforms operating under voluntary data collection, such as Spotify or Netflix, must adapt their recommendation systems based on users' privacy decisions, choosing between personalized recommendations (based on data available) and generalized recommendations (based on data unavailable). To analyze the strategic interactions between users and platforms. Solving it via backward induction, given the platform's recommendation portfolio (**Step 1**), by verifying the user's privacy disclosure decision and activity decision (**Step 2**), we can obtain the final equilibrium (**Step 3**).

3.4.1 Step 1: The platform's recommendation portfolio

The platform would make his recommendation portfolio prior to the user's privacy choice. Based on the history data availability, we divide the portfolio into two parts, i.e., personalize recommendation strategy (in terms of data available) and generalize recommendation strategy (in terms of data unavailable), as shown in Table 3.4.

Table 3.4: Recommendation portfolio in Voluntary Collection Scenario

	Data available	Data unavailable
Portfolio 1	Polarize	i
Portfolio 2	Polarize	j
Portfolio 3	Diversity	i
Portfolio 4	Diversity	j
Portfolio 5	Generalize i	i
Portfolio 6	Generalize i	j
Portfolio 7	Generalize j	i
Portfolio 8	Generalize j	j

3.4.2 Step 2: Users' privacy disclosing decision and activity decision

Different preference types of users with heterogeneous data history can result in their privacy decisions (disclose data or not) by anticipating the platform's recommendation portfolio that can maximizing their $u^k_{\langle H,P\rangle}$.

Part (a): We will show the utility maximization conditions for different users, i.e., the user's preference type:

(1) For the the $\langle i, i\rangle$ user, one can verify that:

  (a) $u^i_{\langle i,i\rangle} < u^j_{\langle i,i\rangle}$ holds for $\theta \in [0, \bar{\theta}_V]$, she would prefer product j;

  (b) $u^i_{\langle i,i\rangle} > u^j_{\langle i,i\rangle}$ holds for $\theta \in [\bar{\theta}_V, 1]$, she would prefer product i.

(2) For the the $\langle j, i\rangle$ user, one can verify that $u^i_{\langle j,i\rangle} > u^j_{\langle j,i\rangle}$ holds for any $\theta \in [0, 1]$, she would always prefer product i;

(3) For the the $\langle i, j\rangle$ user, one can verify that $u^i_{\langle i,j\rangle} < u^j_{\langle i,j\rangle}$ holds for any $\theta \in [0, 1]$, she would always prefer product j;

(4) For the the $\langle j, j\rangle$ user, one can verify that:

(a) $u^i_{\langle j,j\rangle} > u^j_{\langle j,j\rangle}$ holds for $\theta \in [0, \bar{\theta}_V]$, she would prefer product i;

(b) $u^i_{\langle j,j\rangle} < u^j_{\langle j,j\rangle}$ holds for $\theta \in [\bar{\theta}_V, 1]$, she would prefer product j.

Note that $\bar{\theta}_V = \frac{(1-\delta)v}{s}$

Part (b): By anticipating the platform's recommendation portfolio, users will choose disclose data or not based on whether their preference type will match the platform's recommendations. We solve users' privacy disclosing decisions as the following (**summarized in Table 3.5 and Table 3.6**):

(1) Recommendation Portfolio 1: given that the platform will certainly recommend the same product to history data available users, and recommend product i to history data unavailable users, one can verify that:

(i) For $\langle i, i\rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation i would always mismatch their preference type j;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation i would always match their preference type i.

(ii) For $\langle j, i\rangle$ users: (Not to disclose) They choose not to disclose data, as the generalized recommendation i would always match their preference type i;

Table 3.5: Privacy decision and recommendation type under different recommendation portfolio

		Portfolio 1		**Portfolio 2**		**Portfolio 3**		**Portfolio 4**	
User	Type	Data	Product	Data	Product	Data	Product	Data	Product
$\langle i,i\rangle$	$0<\theta<\bar{\theta}_V$	Indifferent	i	Unavailable	j	Available	j	Indifferent	j
	$\bar{\theta}_V<\theta<1$	Indifferent	i	Available	i	Unavailable	i	Indifferent	j
$\langle j,i\rangle$	$0<\theta<1$	Unavailable	i	Indifferent	j	Indifferent	i	Available	i
$\langle i,j\rangle$	$0<\theta<1$	Indifferent	i	Unavailable	j	Available	j	Indifferent	j
$\langle j,j\rangle$	$0<\theta<\bar{\theta}_V$	Unavailable	i	Indifferent	j	Indifferent	i	Available	i
	$\bar{\theta}_V<\theta<1$	Available	j	Indifferent	j	Indifferent	i	Unavailable	j

*Note that “Indifferent” refers to the data availability and unavailability does not affect the recommendation results.

Table 3.6: Privacy decision and recommendation type under different recommendation portfolio (Cont.)

		Portfolio 5		Portfolio 6		Portfolio 7		Portfolio 8	
User	Type	Data	Product	Data	Product	Data	Product	Data	Product
$\langle i, i\rangle$	$0 < \theta < \bar{\theta}_V$	Indifferent	i	Unavailable	j	Available	j	Indifferent	j
	$\bar{\theta}_V < \theta < 1$	Indifferent	i	Available	i	Unavailable	i	Indifferent	j
$\langle j, i\rangle$	$0 < \theta < 1$	Indifferent	i	Available	i	Unavailable	i	Indifferent	j
$\langle i, j\rangle$	$0 < \theta < 1$	Indifferent	i	Unavailable	j	Available	j	Indifferent	j
$\langle j, j\rangle$	$0 < \theta < \bar{\theta}_V$	Indifferent	i	Available	i	Unavailable	i	Indifferent	j
	$\bar{\theta}_V < \theta < 1$	Indifferent	i	Unavailable	j	Available	j	Indifferent	j

*Note that “Indifferent” refers to the data availability and unavailability does not affect the recommendation results.

(iii) For $\langle i, j\rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the recommendation i would always mismatch their preference type j;

(iv) For $\langle j, j\rangle$ users:

a. (Not to disclose) They choose not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the generalized recommendation i would always match their preference type i;

b. (Disclose) They choose to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the personalized recommendation j would always match their preference type j;

(2) <u>Recommendation Portfolio 2</u>: given that the platform will certainly recommend the same product to history data available users, and recommend product j to history data unavailable users, one can verify that:

(i) For $\langle i, i\rangle$ users:

a. (Not to disclose) They choose not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the generalized recommendation j would always match their preference type j;

b. (Disclose) They choose disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the personalized recommendation i would always match their preference type i.

(ii) For $\langle j, i\rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the generalized recommendation j would always mismatch their preference type i;

(iii) For $\langle i, j \rangle$ users: (Not to disclose) They choose not to disclose data, as the generalized recommendation j would always match their preference type j;

(iv) For $\langle j, j \rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation j would always mismatch their preference type i;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation j would always match their preference type j;

(3) <u>Recommendation Portfolio 3</u>: given that the platform will certainly recommend the opposite product to history data available users, and recommend product i to history data unavailable users, one can verify that:

(i) For $\langle i, i \rangle$ users:

a. (Disclose) They choose disclose data for $\theta \in [0, \bar{\theta}_V]$, as the personalized recommendation j would always match their preference type j;

b. (Not to Disclose) They choose not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the generalized recommendation i would always mismatch their preference type i.

(ii) For $\langle j, i \rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the recommendation i would always match their preference type i;

(iii) For $\langle i, j \rangle$ users: (Disclose) They choose disclose data, as the personalized recommendation j would always match their preference type j;

(iv) For $\langle j, j \rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation i would always match their preference type i;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation i would always mismatch their preference type j;

(4) <u>Recommendation Portfolio 4</u>: given that the platform will certainly recommend the opposite product to to history data available users, and recommend product j to history data unavailable users, one can verify that:

(i) For $\langle i, i \rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation j would always match their preference type j;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation j would always mismatch their preference type i.

(ii) For $\langle j, i \rangle$ users: (Disclose) They choose disclose data, as the personalized recommendation i would always match their preference type i;

(iii) For $\langle i, j \rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the recommendation j would always match their preference type j;

(iv) For $\langle j, j \rangle$ users:

a. (Disclose) They choose disclose data for $\theta \in [0, \bar{\theta}_V]$, as the personalized recommendation i would always match their preference type i;

b. (Not to disclose) They choose not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the generalized recommendation j would always match their preference type j;

(5) <u>Recommendation Portfolio 5</u>: given that the platform will certainly recommend the personalized product i to to history data available users, and recommend generalized product i to history data unavailable users, one can verify that:

(i) For $\langle i, i \rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation i would always mismatch their preference type j;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation i would always match their preference type i.

(ii) For $\langle j, i \rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the recommendation i would always match their preference type i;

(iii) For $\langle i, j \rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the recommendation i would always mismatch their preference type j;

(iv) For $\langle j, j\rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation i would always match their preference type i;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation i would always mismatch their preference type j;

(6) <u>Recommendation Portfolio 6</u>: given that the platform will certainly recommend the personalized product i to to history data available users, and recommend the generalized product j to history data unavailable users, one can verify that:

(i) For $\langle i, i\rangle$ users:

a. (Not to disclose) They choose not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the generalized recommendation j would always match their preference type j;

b. (Disclose) They choose to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the personalized recommendation i would always match their preference type i.

(ii) For $\langle j, i\rangle$ users: (Disclose) They choose disclose data, as the personalized recommendation i would always match their preference type i;

(iii) For $\langle i, j\rangle$ users: (Not to disclose) They choose not to disclose data, as the generalized recommendation j would always match their preference type j;

(iv) For $\langle j, j\rangle$ users:

a. (Disclose) They choose disclose data for $\theta \in [0, \bar{\theta}_V]$, as the personalized recommendation i would always match their preference type i;

b. (Not to disclose) They choose not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the generalized recommendation j would always match their preference type j;

(7) Recommendation Portfolio 7: given that the platform will certainly recommend the personalized product j to to history data available users, and recommend the generalized product i to history data unavailable users, one can verify that:

(i) For $\langle i, i\rangle$ users:

a. (Disclose) They choose to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the personalized recommendation j would always match their preference type j;

b. (Not to disclose) They choose not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the generalized recommendation i would always match their preference type i.

(ii) For $\langle j, i\rangle$ users: (Not to disclose) They choose not to disclose data, as the generalized recommendation i would always match their preference type i;

(iii) For $\langle i, j \rangle$ users: (Disclose) They choose to disclose data, as the personalized recommendation j would always match their preference type j;

(iv) For $\langle j, j \rangle$ users:

a. (Not to disclose) They choose not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the generalized recommendation i would always match their preference type i;

b. (Disclose) They choose not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the personalized recommendation j would always match their preference type j;

(8) Recommendation Portfolio 8: given that the platform will certainly recommend the personalized product j to to history data available users, and recommend generalized product j to history data unavailable users, one can verify that:

(i) For $\langle i, i \rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation j would always match their preference type j;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation j would always mismatch their preference type i.

(ii) For $\langle j, i \rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the recommendation j would always mismatch their preference type i;

(iii) For $\langle i, j\rangle$ users: (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data, as the recommendation j would always match their preference type j;

(iv) For $\langle j, j\rangle$ users:

a. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [0, \bar{\theta}_V]$, as the recommendation j would always mismatch their preference type i;

b. (Disclose & Not to disclose) They are indifferent between disclose and not to disclose data for $\theta \in [\bar{\theta}_V, 1]$, as the recommendation j would always match their preference type j;

3.4.3 Step 3: The platform's equilibrium results (recommendation decision)

The objective of the platform is a function that is maximizing users' activity $d^k_{\langle H,P\rangle}$ (demand). The platform's expected profit based on personalized and generalized recommendation can be specified in different data availability conditions. Given that users are divided by threshold $\bar{\theta}_V$ in Step 2, in what follows, we proceed to prove parts (a)-(b) below (as shown in Figure **??**).

Part (a). The total surplus and profits in each portfolio can be defined by $S^{(N)} = \gamma S^{(N)}_{\langle H,i\rangle} + (1-\gamma) S^{(N)}_{\langle H,j\rangle}$ and $\pi^{(N)} = \gamma D^{(N)}_{\langle H,i\rangle} + (1-\gamma) D^{(N)}_{\langle H,j\rangle}$, respectively. Specifically,

- $S^{(N)}_{\langle H,i\rangle} = \underbrace{\alpha \left(\int_0^{\bar{\theta}_V} u^k_{\langle i,i\rangle}\, d\theta + \int_{\bar{\theta}_V}^{\left[d^k_{\langle i,i\rangle} \wedge 1\right]} u^k_{\langle i,i\rangle}\, d\theta \right)}_{\text{surplus of } \langle i,i\rangle \text{ users}} + \underbrace{(1-\alpha) \left(\int_0^{\left[\bar{\theta}^k_{\langle j,i\rangle} \wedge 1\right]} u^k_{\langle j,i\rangle}\, d\theta \right)}_{\text{surplus of } \langle j,i\rangle \text{ users}}$

- $S^{(N)}_{\langle H,j\rangle} = \underbrace{\alpha \left(\int_0^{\left[\bar{\theta}^k_{\langle j,i\rangle} \wedge 1\right]} u^k_{\langle i,j\rangle} \, d\theta \right)}_{\text{surplus of } \langle i,j\rangle \text{ users}} + \underbrace{(1-\alpha) \left(\int_0^{\bar{\theta}_V} u^k_{\langle j,j\rangle} \, d\theta + \int_{\bar{\theta}_V}^{\left[d^k_{\langle j,j\rangle} \wedge 1\right]} u^k_{\langle j,j\rangle} \, d\theta \right)}_{\text{surplus of } \langle j,j\rangle \text{ users}}$
- $D^{(N)}_{\langle H,i\rangle} = \underbrace{\alpha \left[(d^k_{\langle i,i\rangle} \wedge \bar{\theta}_V) + (d^k_{\langle i,i\rangle} - \bar{\theta}_V) \wedge (1-\bar{\theta}_V) \right]}_{\text{demand from } \langle i,i\rangle \text{ users}} + \underbrace{(1-\alpha) d^k_{\langle j,i\rangle}}_{\text{demand from } \langle j,i\rangle \text{ users}}$
- $D^{(N)}_{\langle H,j\rangle} = \underbrace{\alpha d^k_{\langle i,j\rangle}}_{\text{demand from } \langle i,j\rangle \text{ users}} + \underbrace{(1-\alpha) \left[(d^k_{\langle j,j\rangle} \wedge \bar{\theta}_V) + (d^k_{\langle j,j\rangle} - \bar{\theta}_V) \wedge (1-\bar{\theta}_V) \right]}_{\text{demand from } \langle j,j\rangle \text{ users}}$

Note that the superscript (N) denote Portfolio N, and $x \wedge y = \min\{x, y\}$

Part (b). In the final equilibrium, one can verify the optimal surplus and profits can be defined by

$$\begin{aligned} \hat{\pi} &= \pi^{\text{Portfolio 6}} = \pi^{\text{Portfolio 7}} = 1 \\ \hat{S} &= S^{\text{Portfolio 6}} = S^{\text{Portfolio 7}} \end{aligned} \tag{3.7}$$

1. **Leveraging User data control right to maximize demand and Profit**: The superior profit and consumer surplus in Scenario $\mathbb{V}$ underscore the value of granting users control over their data disclosure. By allowing users to voluntarily share historical data, platforms can align recommendations more closely with individual preferences, enhancing user satisfaction and engagement. For instance, social media platforms like TikTok can implement opt-in data-sharing prompts that highlight personalized content benefits (e.g., tailored “For You” feeds), encouraging users to disclose data. This user-centric approach not only boosts consumer surplus by delivering relevant content but also maximizes platform profit through increased user activity and ad revenue. Managers should invest in transparent data consent interfaces to build trust, ensuring users perceive data sharing as a value-

enhancing choice.

2. **Personalized recommendation portfolios for heterogeneous users**: The success of Portfolios 6 and 7 in Scenario $\mathbb{V}$ highlights the efficacy of hybrid recommendation strategies that combine personalized recommendations (for users disclosing data) and generalized recommendations (for users opting out). Managers can operationalize this by segmenting users based on their privacy decisions and tailoring recommendations accordingly. For example, an e-commerce platform like Amazon could offer personalized product suggestions to users who share browsing history while providing generalized category-based recommendations (e.g., "Top Picks in Electronics") to those who opt out. By dynamically adjusting recommendation algorithms to accommodate both user types, platforms can capture maximum demand across diverse user segments, reinforcing network effects and sustaining long-term profitability.

3. **Highlight win-win situation**: The alignment of high platform profit and consumer surplus in Scenario $\mathbb{V}$ suggests a win-win scenario that managers can capitalize on through strategic communication. By clearly articulating the mutual benefits of voluntary data sharing—enhanced user experience for consumers and improved service offerings for the platform—managers can increase user participation rates. For instance, a platform could launch educational campaigns illustrating how data-driven personalization improves content relevance (e.g., Spotify's "Discover Weekly" playlists) while emphasizing robust data protection measures. Such communication fosters a positive feedback loop, where increased user trust drives higher data disclo-

sure, further amplifying profit and surplus.

Chapter 4

Impact of Data Policy on Platform Service Level

4.1 Model Framework

Consider a platform (*he*) hosting two types of products: *polarized* products (a) and *diversified* products (b). He chooses a data-driven product recommendation strategy that specifies (i) *a data utilization strategy* – the platform decides to what extent to collect/learn from the user's data; (ii) *a recommendation strategy* – trade-offs in recommending different products to users. Each user (*she*) decides how to respond to the platform's product recommendations – stay in the platform or not. By analyzing the users' data, the platform learns about relevant users' characteristics and provides higher-matching recommendations to them. We next introduce formally these elements.

4.1.1 Users

Each user decides how to respond to the platform's product recommendations. Their utility consists of two parts: Part (1) illustrates her product valuation and Part (2) illustrates her disutility.

Part (1): product valuation. Users' base valuation to product $i, i \in \{a, b\}$ is $v = \mathbb{I}_{\text{like}}$. However, they different in following ways: (i) regarding to the polarized product a, users are heterogeneous in their polarization degree $\delta, \delta \sim \text{Uniform}[0, 1]$, their valuation to product a can be defined as $\delta\mathbb{I}_{\text{like}}$; (ii) regarding to the diversified product b, before users searching for the unknown, the prior probability of diversified product matching their preference can be defined as $P\{v = 1\} = \frac{1}{2}$ and $P\{v = 0\} = \frac{1}{2}$.

Part (2): disutility. *Firstly*, Each users will incur an attention fee p when they accept a product[1]. *Secondly*, when a user receives a product that is opposite to her preference, she will either incur a searching cost s to searching for diversified product b, or freely back to the polarized product a, otherwise, users will leave the recommendation process. Figure 4.1 summarizes the consumer's choices graphically. Note that the subscript and superscript in user's utility $u^k_{\langle i,-i\rangle}$ represents $\langle$Platform's recommendation, The user's choice$\rangle$ and $k = \{$accept, search, reject$\}$, respectively. Further, we make the following assumption about s:

ASSUMPTION 1. $s < \mathbb{E}[v]$.

Assumption 1 is typical because: (i) it excludes trivial cases where users failing to derive positive utility from searching another products; (ii) it also turns

[1] Attention fee refer to the time and energy users invest in a platform, often unconsciously, in exchange for "free" services (e.g., browsing Instagram means users pay with their attention, while the platform earns money through advertising).

users' second search unlikely to continue indefinitely because it's costly for users to spending their limited time, i.e., users will search for only one time.

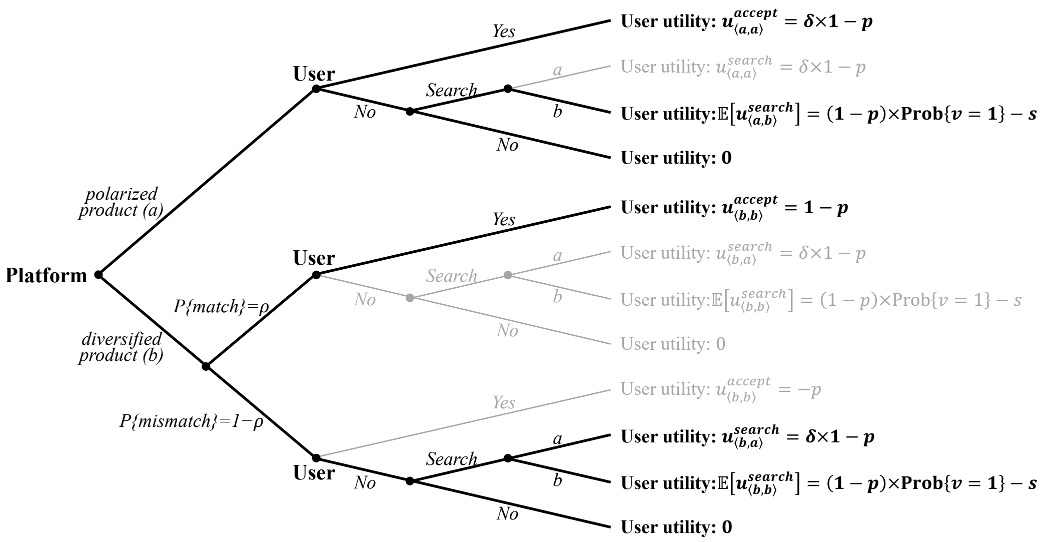


Figure 4.1: Game tree

4.1.2 The platform

The platform's profit maximization is dependent upon the *data utilization strategy* and the *recommendation strategy*. (1) Firstly, regarding the data utilization strategy: on the one hand, the paper considers the *Data policy* that specifies between three schemes: (i) Mandatory collection (denoted by scenario $\mathbb{M}$); (ii) Data free collection (denoted by scenario $\mathbb{N}$); (iii) Voluntary collection (denoted by scenario $\mathbb{V}$). The key difference between the three schemes (denoted by superscript $k, k \in \{\mathbb{N}, \mathbb{M}, \mathbb{V}\}$) is whether the platform has the data to learn users' polarization level δ and match user preference with probability ρ. On the other hand, when recommending diversified product b, the platform can use data to better match user's preferences, denoted by $\mathrm{P}\{\text{match}\} = \rho^k, \rho^k \in [0, 1]$; on the other hand, when recommending polarized product, the platform can learn user's polarization preference δ.

(2) Secondly, regarding the recommendation strategies: We categorize the platform's recommendation into two schemes: (i) *Generalized recommendation* – The platform will recommend a single product (either a or b) to users whose data unavailable, depending on which product brings him the highest demand. (ii) *Personalized recommendation* – the platform recommend products that meets user preference if her data available. As such, the platform's total profit is given by $\pi = p(d_a + d_b)$, where d_i denote the platform's demand of product i, for $i \in \{a, b\}$.

4.1.3 Sequence of Events

There are two stages in our model, with detailed game sequences given below.

Stage 1 — Recommendation and pricing decision: Platform decides: (i) recommend which types of product to users. (ii) product price. Without loss of generality, we use the following tie-breaking rule: if the platform is indifferent between recommending product a and b, he will recommending product a.

Stage 2 — Purchasing decision: Given the platform's recommendation strategy, users could choose (i) accept current product recommendation; (ii) searching for the opposite product; (iii) leave the recommendation system.

4.2 Data-free Collection Scenario ($\mathbb{N}$)

Under $\mathbb{N}$ model, the platform making his generalized recommendation decision as he cannot distinguish each users, i.e., each user receives the same product type, either a or b.

Following the backward induction approach and use "*" to denote the final equilibrium outcome, we complete the equilibrium in two steps: **Step 1.** Solving users' purchasing decision. **Step 2.** Solving the platform's recommendation and pricing strategy.

Step (1). Users' purchasing decision.

Before solving users' purchasing decision to product i, one should verify that: given the uncertainty of users' utility to product b, if mismatch happens, their utility of searching b can be either positive (i.e., $u^{\text{search}}_{\langle i,b\rangle} > 0$, which is equivalent to $p < \bar{p}$) or negative, (i.e., $u^{\text{search}}_{\langle i,b\rangle} \leqslant 0$, which is equivalent to $p \geqslant \bar{p}$). Furthermore, we specifying four subcases to one particular user:

- Subcase (i). The user was recommended by product a and $u^{\text{search}}_{\langle i,b\rangle} > 0$. It follows that this user will:

$$\begin{cases} \text{Choose product } a, & \text{if } \delta \in [\bar{\delta}, 1]; \\ \text{Search product } b, & \text{if } \delta \in [0, \bar{\delta}] \end{cases}$$

- Subcase (ii). The user was recommended by product a and $u^{\text{search}}_{\langle i,b\rangle} \leqslant 0$. It follows that this user will:

$$\begin{cases} \text{Choose product } a, & \text{if } \delta \in [p, 1]; \\ \text{Leave the market}, & \text{if } \delta \in [0, p] \end{cases}$$

- Subcase (iii). The user was recommended by product b and $u^{\text{search}}_{\langle i,b\rangle} > 0$. It follows that this user will:

$$\begin{cases} \text{With P}\{\text{match}\}\text{=}\rho\text{: Choose product } b \\ \text{With P}\{\text{mismatch}\}\text{=}1-\rho: \begin{cases} \text{Searching for product } a, & \text{if } \delta \in [\bar{\delta}, 1]; \\ \text{Searching for product } b, & \text{if } \delta \in [0, \bar{\delta}] \end{cases} \end{cases}$$

- Subcase (iv). The user was recommended by product b and $u^{\text{search}}_{\langle i,b\rangle} \leqslant 0$. It follows that this user will:

$$\begin{cases} \text{With P}\{\text{match}\}\text{=}\rho\text{: Choose product } b \\ \text{With P}\{\text{mismatch}\}\text{=}1-\rho: \begin{cases} \text{Searching for product } a, & \text{if } \delta \in [p, 1]; \\ \text{Leave the market}, & \text{if } \delta \in [0, p] \end{cases} \end{cases}$$

(Note: $\bar{p} = 1 - 2s, \bar{\delta} = \frac{1+p-2s}{2}$ and $\bar{\delta} \in [0, 1]$ holds.)

Step (2). Platform's recommendation and pricing decision.

Part (i). Recommendation strategy. To facilitate our derivation, we first process the recommendation strategy given the pricing scheme. We categorize two pricing scheme:

- Users searching b with positive utility, i.e., $p < \bar{p}$.

$$\begin{cases} \text{If recommend product } a, & \begin{cases} d_a = 1 - \bar{\delta}; \\ d_b = \bar{\delta}/2 \end{cases} \\ \text{If recommend product } b, & \begin{cases} d_b = \rho^N + (1-\rho^N)\bar{\delta}/2; \\ d_a = (1-\rho^N)(1-\bar{\delta}) \end{cases} \end{cases}$$

By comparing π_a and π_b in Part (i) - case (a), we can know that $\pi_a \leqslant \pi_b$, resulting the platform's generalized recommendation strategy: recommending product b.

- Users searching b with negative utility, i.e., $p \geqslant \bar{p}$.

$$\begin{cases} \text{If recommend product } a, & \begin{cases} d_a = 1 - p; \\ d_b = 0 \end{cases} \\ \text{If recommend product } b, & \begin{cases} d_b = \rho^N; \\ d_a = (1 - \rho^N)(1 - p) \end{cases} \end{cases}$$

By comparing π_a and π_b in Part (i) - case (b), we can know that $\pi_a \leqslant \pi_b$, resulting the platform's generalized recommendation strategy: <u>recommending product b</u>.

Part (ii). Pricing strategy.

Firstly, based on the recommendation portfolio, we can rewrite the profit function as the following:

$$\pi^N = \begin{cases} p[\rho^N + (1 - \rho^N)(1 - \bar{\delta} + \bar{\delta}/2)], & \text{if } p < \bar{p}, \quad \text{Recommending } b \\ p[\rho^N + (1 - \rho^N)(1 - p)], & \text{if } p \geqslant \bar{p}, \quad \text{Recommending } b \end{cases}$$

Secondly, solving π^N in each segments, we can know that:

- If $p < \bar{p}$, the platform's profit from recommending product b is strictly increasing in p, since that $\frac{\partial \pi^N}{\partial p} > 0$ if $p < \bar{p}$. Thus, it leads to the local optimal price of $\hat{p}^N = \bar{p}$ and $\pi^N = \frac{1}{2}\left[\rho^N + 4(\rho^N - 1)s^2 - 4\rho^N s + 1\right]$

- If $p \geqslant \bar{p}$, the platform's profit from recommending product b is concave in p, i.e., $\frac{\partial^2 \pi^N}{\partial p^2} < 0$. Thus, solving the first-order-condition leads to the optimal price of $\hat{p}^N = \frac{1}{2(1-\rho^N)}$. We distinguish two cases:

  1. Case (i): $\hat{p}^N \geqslant 1$, which is equivalent to $\rho^N \in [\frac{1}{2}, 1]$, it follows that

the platform's profit from recommending product b maximize at $\hat{p} = 1$ and the resulting profit is $\hat{\pi}^N = \rho^N$;

2. Case (ii): $\hat{p}^N < 1$, which is equivalent to $\rho^N \in [0, \frac{1}{2})$, it follows that the local optimal results are:

$$\hat{p}^N = \begin{cases} \frac{1}{2(1-\rho^N)}, & \text{if } s \in \left[\bar{s}_1^N, \frac{1}{2}\right] \\ 1-2s, & \text{if } s \in \left[0, \bar{s}_1^N\right] \end{cases}$$

$$\hat{\pi}^N = \begin{cases} \frac{1}{4(1-\rho^N)}, & \text{if } s \in \left[\bar{s}_1^N, \frac{1}{2}\right] \\ (1-2s)[\rho^N - 2(\rho^N - 1)s], & \text{if } s \in \left[0, \bar{s}_1^N\right] \end{cases}$$

Part (iii). The final equilibrium. The platform choose to recommend product b to all users. Furthermore, We should find the global optimal results as the following (illustrated in Figure **??**):

- If $\rho^N \in \left[\frac{1}{2}, 1\right]$, then we have:

$$p^{N*} = \begin{cases} 1, & \text{if } s \in [\bar{s}_2^N, \frac{1}{2}]; \\ 1-2s, & \text{if } s \in [0, \bar{s}_2^N] \end{cases}$$

$$D^{N*} = \begin{cases} d_a = (1-\rho^N)(1-p^{N*}), \quad \text{and } d_b = \rho^N, & \text{if } s \in [\bar{s}_2^N, \frac{1}{2}]; \\ d_a = (1-\rho^N)(1-\bar{\delta}), \quad \text{and } d_b = \rho^N + (1-\rho^N)\bar{\delta}/2, & \text{if } s \in [0, \bar{s}_2^N] \end{cases}$$

$$\pi^{N*} = \begin{cases} \rho^N, & \text{if } s \in [\bar{s}_2^N, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^N + 4(\rho^N - 1)s^2 - 4\rho^N s + 1\right], & \text{if } s \in [0, \bar{s}_2^N] \end{cases}$$

$$CS^{N*} = \begin{cases} \rho^N(1-p^{N*}) + (1-\rho^N)\int_{p^{N*}}^1 (\delta - p^{N*})\, d\delta, & \text{if } s \in [\bar{s}_2^N, \frac{1}{2}]; \\ \rho^N(1-p^{N*}) + (1-\rho^N)\left[\int_{\bar{\delta}}^1 (\delta - p^{N*})\, d\delta + \frac{\bar{\delta}}{2}(1-p^{N*})\right], & \text{if } s \in [0, \bar{s}_2^N] \end{cases}$$

$$= \begin{cases} 0, & \text{if } s \in [\bar{s}_2^N, \frac{1}{2}]; \\ s(\rho^N + 1), & \text{if } s \in [0, \bar{s}_2^N] \end{cases}$$

- If $\rho^N \in \left[0, \frac{1}{2}\right)$, then we have:

$$p^{N*} = \begin{cases} \frac{1}{2(1-\rho)}, & \text{if } s \in [\bar{s}_3^N, \frac{1}{2}]; \\ 1-2s, & \text{if } s \in [\bar{s}_1^N, \bar{s}_3^N] \\ 1-2s, & \text{if } s \in [0, \bar{s}_1^N] \end{cases}$$

$$D^{N*} = \begin{cases} d_a = (1-\rho^N)(1-p^{N*}), & \text{and } d_b = \rho^N, & \text{if } s \in [\bar{s}_3^N, \frac{1}{2}]; \\ d_a = (1-\rho^N)(1-\bar{\delta}), & \text{and } d_b = \rho^N + (1-\rho^N)\bar{\delta}/2, & \text{if } s \in [\bar{s}_1^N, \bar{s}_3^N] \\ d_a = (1-\rho^N)(1-\bar{\delta}), & \text{and } d_b = \rho^N + (1-\rho^N)\bar{\delta}/2, & \text{if } s \in [0, \bar{s}_1^N] \end{cases}$$

$$\pi^{N*} = \begin{cases} \frac{1}{4(1-\rho^N)}, & \text{if } s \in [\bar{s}_3^N, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^N + 4(\rho^N-1)s^2 - 4\rho s + 1\right], & \text{if } s \in [\bar{s}_1^N, \bar{s}_3^N] \\ \frac{1}{2}\left[\rho^N + 4(\rho^N-1)s^2 - 4\rho s + 1\right], & \text{if } s \in [0, \bar{s}_1^N] \end{cases}$$

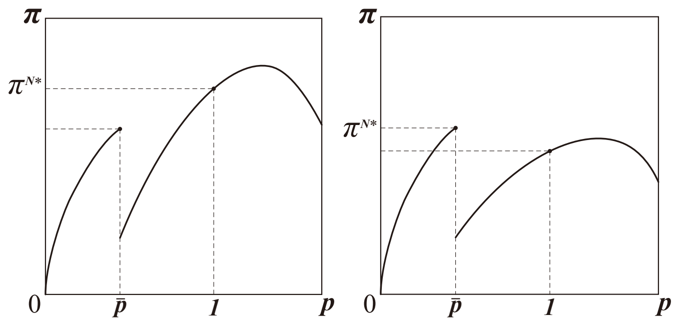
$$CS^{N*} = \begin{cases} \rho^N(1-p^{N*}) + (1-\rho^N)\int_{p^{N*}}^1 (\delta - p^{N*})\, d\delta, & \text{if } s \in [\bar{s}_3^N, \frac{1}{2}]; \\ \rho^N(1-p^{N*}) + (1-\rho^N)\left[\int_{\bar{\delta}}^1 (\delta - p^{N*})\, d\delta + \frac{\bar{\delta}}{2}(1-p^{N*})\right], & \text{if } s \in [\bar{s}_1^N, \bar{s}_3^N] \\ \rho^N(1-p^{N*}) + (1-\rho^N)\left[\int_{\bar{\delta}}^1 (\delta - p^{N*})\, d\delta + \frac{\bar{\delta}}{2}(1-p^{N*})\right], & \text{if } s \in [0, \bar{s}_1^N] \end{cases}$$

$$= \begin{cases} \frac{1-4(\rho^N)^2}{8-8\rho^N}, & \text{if } s \in [\bar{s}_3^N, \frac{1}{2}]; \\ s(\rho^N+1), & \text{if } s \in [\bar{s}_1^N, \bar{s}_3^N] \\ s(\rho^N+1) & \text{if } s \in [0, \bar{s}_1^N] \end{cases}$$

Note that $\bar{s}_1^N = \frac{2\rho^N-1}{4\rho^N-4}$, $\bar{s}_2^N = \frac{\rho^N - \sqrt{2(\rho^N-1)\rho^N+1}}{2(\rho^N-1)}$, $\bar{s}_3^N = \frac{\sqrt{2}-2\rho^N}{4-4\rho^N}$ and $\bar{s}_1^N < \bar{s}_3^N$ holds.

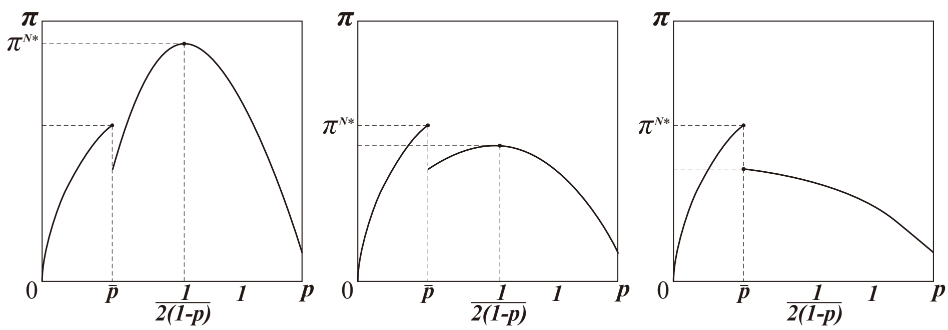
4.3 Mandatory Collection Scenario ($\mathbb{M}$)

Under $\mathbb{M}$ model, the platform making his personalized recommendation decision, each users are recommended with different products depending on her polarization preference, i.e., the market is partitioned by the platform. Following the backward



(a) $s \in [\bar{s}_2, 1]$ (b) $s \in [0, \bar{s}_2)$

Figure 4.2: Categories of global optimal results in Scenario $\mathbb{N}$ if $\rho \in [\frac{1}{2}, 1]$



(a) $s \in [\bar{s}_3, \frac{1}{2}]$ (b) $s \in [\bar{s}_1, \bar{s}_3)$ (c) $\rho^k \in [0, \bar{s}_1)$

Figure 4.3: Categories of global optimal results in Scenario $\mathbb{N}$ if $\rho \in [0, \frac{1}{2}]$

induction approach, we complete the equilibrium in two steps: **Step 1.** Solving users' purchasing decision. **Step 2.** Solving the platform's recommendation and pricing strategy.

Step (1). Users' purchasing decision.

The user's decision is the same as Scenario $\mathbb{N}$, so we omitted step (1) here.

Step (2). Platform's recommendation and pricing decision.

Part (i). Recommendation strategy. To facilitate our derivation, we first process the recommendation strategy given the pricing scheme. We categorize two pricing scheme:

- Users searching b with positive utility, i.e., $p < \bar{p}$.

$$\begin{cases} \text{For users located in } \delta \in [\bar{\delta}, 1], & \begin{cases} \text{If recommend product } a, & \begin{cases} d_a = (1-\bar{\delta}) \\ d_b = 0 \end{cases} \\ \text{If recommend product } b, & \begin{cases} d_a = (1-\rho^M)(1-\bar{\delta}) \\ d_b = \rho^M(1-\bar{\delta}) \end{cases} \end{cases} \\ \text{For users located in } \delta \in [0, \bar{\delta}], & \begin{cases} \text{If recommend product } a, & \begin{cases} d_a = 0 \\ d_b = \bar{\delta}/2 \end{cases} \\ \text{If recommend product } b, & \begin{cases} d_a = 0 \\ d_b = \rho^M\bar{\delta} + (1-\rho^M)\bar{\delta}/2 \end{cases} \end{cases} \end{cases}$$

By comparing the profits in each market segments with different recommendation, we can show the recommendation strategy as the following:

$$\begin{cases} \text{For users located in } \delta \in [\bar{\delta}, 1], & \text{recommending product } a \\ \text{For users located in } \delta \in [0, \bar{\delta}], & \text{recommending product } b \end{cases}$$

- Users searching b with negative utility, i.e., $p \geqslant \bar{p}$.

$$\begin{cases} \text{For users located in } \delta \in [p,1], & \begin{cases} \text{If recommend product } a, & \begin{cases} d_a = 1-p \\ d_b = 0 \end{cases} \\ \text{If recommend product } b, & \begin{cases} d_a = (1-\rho^M)(1-p) \\ d_b = \rho^M(1-p) \end{cases} \end{cases} \\ \text{For users located in } \delta \in [0,p], & \begin{cases} \text{If recommend product } a, & \begin{cases} d_a = 0 \\ d_b = 0 \end{cases} \\ \text{If recommend product } b, & \begin{cases} d_a = 0 \\ d_b = p\rho^M \end{cases} \end{cases} \end{cases}$$

By comparing the profits in each market segments with different recommendation, we can show the recommendation strategy as the following:

$$\begin{cases} \text{For users located in } \delta \in [p,1], & \text{recommending product } a \\ \text{For users located in } \delta \in [0,p], & \text{recommending product } b \end{cases}$$

Part (ii). Pricing strategy.

Firstly, based on the recommendation portfolio, we can rewrite the profit function as the following:

$$\pi^M = \begin{cases} p\left[(1-\bar{\delta}) + (\rho^M\bar{\delta} + (1-\rho^M)\bar{\delta}/2)\right], & \text{if } p < \bar{p}, \quad \text{recommending } a \text{ in } \delta \in [\bar{\delta},1] \text{ and } b \text{ otherwise} \\ p\left[(1-p) + p\rho^M\right], & \text{if } p \geqslant \bar{p}, \quad \text{recommending } a \text{ in } \delta \in [\bar{\delta},1] \text{ and } b \text{ otherwise} \end{cases}$$

Secondly, solving π^N in each segments, we can know that:

- If $p < \bar{p}$, the platform's profit from recommending product b is strictly increasing in p, since that $\frac{\partial \pi^M}{\partial p} > 0$ if $p < \bar{p}$. Thus, it leads to the local optimal price of $\hat{p}^M = \bar{p}$ and $\pi^N = \frac{1}{2}\left[\rho^M + 4(\rho^M - 1)s^2 - 4\rho^M s + 1\right]$

- If $p \geqslant \bar{p}$, the platform's profit from recommending product b is concave in p, i.e., $\frac{\partial^2 \pi^M}{\partial p^2} < 0$. Thus, solving the first-order-condition leads to the optimal price of $\hat{p}^M = \frac{1}{2(1-\rho)}$. We distinguish two cases:

1. Case (i): $\hat{p}^M \geqslant 1$, which is equivalent to $\rho^M \in [\frac{1}{2}, 1]$, it follows that the platform's profit from recommending product b maximize at $\hat{p} = 1$ and the resulting profit is $\hat{\pi}^M = \rho^M$;
2. Case (ii): $\hat{p}^M < 1$, which is equivalent to $\rho^M \in [0, \frac{1}{2})$, it follows that the local optimal results are:

$$\hat{p}^M = \begin{cases} \frac{1}{2(1-\rho^M)}, & \text{if } s \in \left[\bar{s}_1^M, \frac{1}{2}\right] \\ 1 - 2s, & \text{if } s \in \left[0, \bar{s}_1^M\right] \end{cases}$$
$$\hat{\pi}^M = \begin{cases} \frac{1}{4(1-\rho^M)}, & \text{if } s \in \left[\bar{s}_1^M, \frac{1}{2}\right] \\ (1-2s)(\rho^M - 2(\rho^M - 1)s), & \text{if } s \in \left[0, \bar{s}_1^M\right] \end{cases}$$

Part (iii). The final equilibrium. The platform choose to recommend product b to all users. Furthermore, we should find the global optimal results as the following (illustrated in Figure 4.4-4.5):

- If $\rho^M \in \left[\frac{1}{2}, 1\right]$, then we have:

$$p^{M*} = \begin{cases} 1, & \text{if } s \in [\bar{s}_2^M, \frac{1}{2}]; \\ 1 - 2s, & \text{if } s \in [0, \bar{s}_2^M] \end{cases}$$

$$D^{M*} = \begin{cases} d_a = 1 - p^{M*}, \quad \text{and } d_b = p^{M*}\rho^M, & \text{if } s \in [\bar{s}_2^M, \frac{1}{2}]; \\ d_a = 1 - \bar{\delta}, \quad \text{and } d_b = \rho^M\bar{\delta} + (1 - \rho^M)\bar{\delta}/2, & \text{if } s \in [0, \bar{s}_2^M] \end{cases}$$

$$\pi^{M*} = \begin{cases} \rho^M, & \text{if } s \in [\bar{s}_2^M, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^M + 4(\rho^M - 1)s^2 - 4\rho^M s + 1\right], & \text{if } s \in [0, \bar{s}_2^M] \end{cases}$$

$$CS^{M*} = \begin{cases} p^{M*}\rho^M(1 - p^{M*}) + \int_{p^{M*}}^1 (\delta - p^{N*})\, d\delta, & \text{if } s \in [\bar{s}_2^M, \frac{1}{2}]; \\ \int_{\bar{\delta}}^1 (\delta - p^{M*})\, d\delta + \left[\frac{\bar{\delta}}{2}(1 - \rho^M) + \rho^M\bar{\delta}\right](1 - p^{M*}), & \text{if } s \in [0, \bar{s}_2^M] \end{cases}$$

$$= \begin{cases} 0, & \text{if } s \in [\bar{s}_2^M, \frac{1}{2}]; \\ s(\rho^M - 2\rho^M s + 1), & \text{if } s \in [0, \bar{s}_2^M] \end{cases}$$

- If $\rho^M \in \left[0, \frac{1}{2}\right)$, then we have:

$$p^{M*} = \begin{cases} \frac{1}{2(1-\rho^M)}, & \text{if } s \in [\bar{s}_3^M, \frac{1}{2}]; \\ 1-2s, & \text{if } s \in [\bar{s}_1^M, \bar{s}_3^M] \\ 1-2s, & \text{if } s \in [0, \bar{s}_1^M] \end{cases}$$

$$D^{M*} = \begin{cases} d_a = 1-p^{M*}, & \text{and } d_b = p^{M*}\rho^M, & \text{if } s \in [\bar{s}_3^M, \frac{1}{2}]; \\ d_a = 1-\bar{\delta}, & \text{and } d_b = \rho^M\bar{\delta} + (1-\rho^M)\bar{\delta}/2, & \text{if } s \in [\bar{s}_1^M, \bar{s}_3^M] \\ d_a = 1-\bar{\delta}, & \text{and } d_b = \rho^M\bar{\delta} + (1-\rho^M)\bar{\delta}/2, & \text{if } s \in [0, \bar{s}_1^M] \end{cases}$$

$$\pi^{M*} = \begin{cases} \frac{1}{4(1-\rho^M)}, & \text{if } s \in [\bar{s}_3^M, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^M + 4(\rho^M-1)s^2 - 4\rho s + 1\right], & \text{if } s \in [\bar{s}_1^M, \bar{s}_3^M] \\ \frac{1}{2}\left[\rho^M + 4(\rho^M-1)s^2 - 4\rho s + 1\right], & \text{if } s \in [0, \bar{s}_1^M] \end{cases}$$

$$CS^{M*} = \begin{cases} p^{M*}\rho^M(1-p^{M*}) + \int_{p^{M*}}^1 (\delta - p^{M*})\, d\delta, & \text{if } s \in [\bar{s}_3^M, \frac{1}{2}]; \\ \int_{\bar{\delta}}^1 (\delta - p^{M*})\, d\delta + \left[\frac{\bar{\delta}}{2}(1-\rho^M) + \rho^M\bar{\delta}\right](1-p^{M*}), & \text{if } s \in [\bar{s}_1^M, \bar{s}_3^M] \\ \int_{\bar{\delta}}^1 (\delta - p^{M*})\, d\delta + \left[\frac{\bar{\delta}}{2}(1-\rho^M) + \rho^M\bar{\delta}\right](1-p^{M*}), & \text{if } s \in [0, \bar{s}_1^M] \end{cases}$$

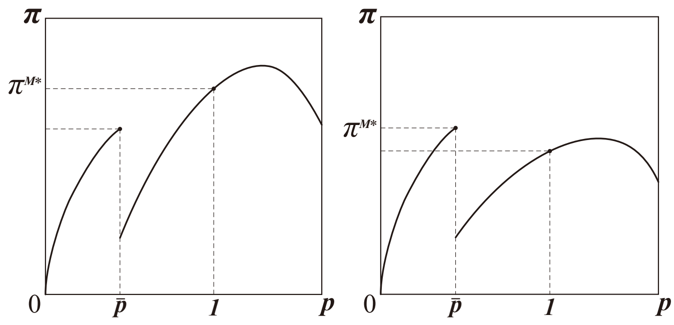
$$= \begin{cases} \frac{1-2\rho^M}{8(\rho^M-1)^2}, & \text{if } s \in [\bar{s}_3^M, \frac{1}{2}]; \\ s(\rho^M - 2\rho^M s + 1), & \text{if } s \in [\bar{s}_1, \bar{s}_3^M] \\ s(\rho^M - 2\rho^M s + 1), & \text{if } s \in [0, \bar{s}_1^M] \end{cases}$$

Note that $\bar{s}_1^M = \frac{2\rho^M-1}{4\rho^M-4}$, $\bar{s}_2^M = \frac{\rho^M - \sqrt{2(\rho^M-1)\rho^M+1}}{2(\rho^M-1)}$, $\bar{s}_3^M = \frac{\sqrt{2}-2\rho^M}{4-4\rho^M}$ and $\bar{s}_1^M < \bar{s}_3^M$ holds.

4.4 Comparison between Scenario $\mathbb{N}$ and $\mathbb{M}$

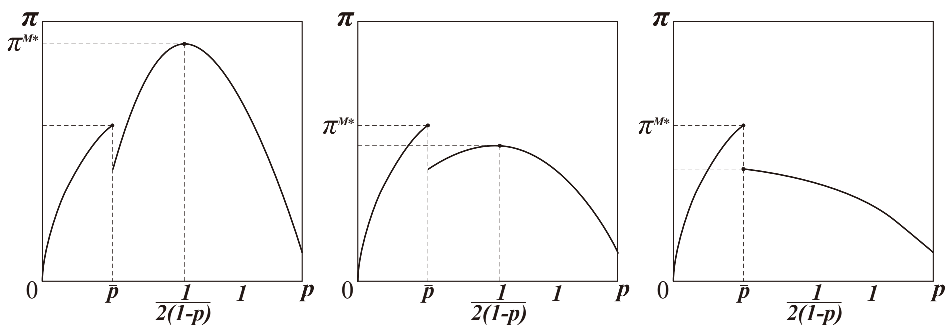
4.4.1 Profits

We categorize three cases:



(a) $s \in [\bar{s}_2, 1]$ (b) $s \in [0, \bar{s}_2)$

Figure 4.4: Categories of global optimal results in Scenario $\mathbb{M}$ if $\rho \in [\frac{1}{2}, 1]$



(a) $s \in [\bar{s}_3, \frac{1}{2}]$ (b) $s \in [\bar{s}_1, \bar{s}_3$ (c) $\rho^k \in [0, \bar{s}_1)$

Figure 4.5: Categories of global optimal results in Scenario $\mathbb{M}$ if $\rho \in [0, \frac{1}{2}]$

- Case (1): $\rho^N = \rho^M$.

  One can easily verify that $\pi^{N*} = \pi^{M*}$ when $\rho^N = \rho^M$. As the total demand remains the same under each scenario.

- Case (2): $\rho^N < \rho^M$. It follows that $0 < \bar{s}_2^M < \bar{s}_2^N < \frac{1}{2}$ and $0 < \bar{s}_3^M < \bar{s}_3^N < \frac{1}{2}$

  Firstly, if $\rho^k \in [\frac{1}{2}, 1]$, then we have:

$$\pi^{N*} = \begin{cases} \rho^N, & \text{if } s \in [\bar{s}_2^N, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^N + 4(\rho^N - 1)s^2 - 4\rho^N s + 1\right], & \text{if } s \in [0, \bar{s}_2^N] \end{cases}$$
$$\pi^{M*} = \begin{cases} \rho^M, & \text{if } s \in [\bar{s}_2^M, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^M + 4(\rho^M - 1)s^2 - 4\rho^M s + 1\right], & \text{if } s \in [0, \bar{s}_2^M] \end{cases}$$

  Secondly, if $\rho^k \in [0, \frac{1}{2}]$, then we have:

$$\pi^{N*} = \begin{cases} \frac{1}{4(1-\rho^N)}, & \text{if } s \in [\bar{s}_3^N, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^N + 4(\rho^N - 1)s^2 - 4\rho s + 1\right], & \text{if } s \in [\bar{s}_1^N, \bar{s}_3^N] \\ \frac{1}{2}\left[\rho^N + 4(\rho^N - 1)s^2 - 4\rho s + 1\right], & \text{if } s \in [0, \bar{s}_1^N] \end{cases}$$
$$\pi^{M*} = \begin{cases} \frac{1}{4(1-\rho^M)}, & \text{if } s \in [\bar{s}_3^M, \frac{1}{2}]; \\ \frac{1}{2}\left[\rho^M + 4(\rho^M - 1)s^2 - 4\rho s + 1\right], & \text{if } s \in [\bar{s}_1^M, \bar{s}_3^M] \\ \frac{1}{2}\left[\rho^M + 4(\rho^M - 1)s^2 - 4\rho s + 1\right], & \text{if } s \in [0, \bar{s}_1^M] \end{cases}$$

  To conclude, by comparison, one can easily verify that $\pi^{N*} \leqslant \pi^{M*}$ holds when $\rho^N < \rho^M$.

- Case (3): $\rho^N > \rho^M$.

  By similar technologies, one can easily verify that $\pi^{N*} \geqslant \pi^{M*}$ holds when $\rho^N > \rho^M$.

1. **Optimizing Recommendation Accuracy Across Scenarios**: The equivalence of profits when $\rho^N = \rho^M$ suggests that, when the probability of matching user preferences is identical in both scenarios, the choice between data-free and mandatory collection does not significantly impact profitability. Managers should focus on ensuring that recommendation algorithms achieve comparable accuracy (ρ) regardless of data collection strategy. For platforms like YouTube, this could involve benchmarking the performance of generalized recommendations in Scenario N (e.g., based on trending content) against personalized recommendations in Scenario M (e.g., based on user history). In markets where $\rho^N = \rho^M$, managers can prioritize the scenario that aligns with operational goals, such as reducing data processing costs in Scenario N or leveraging comprehensive data in Scenario M to enhance user trust and engagement.

2. **Leveraging Mandatory Collection for Superior Recommendation Precision**: The finding that profit in Scenario M is higher when $\rho^N = \rho^M$ under certain conditions indicates that mandatory data collection can yield greater profitability when it enables more precise recommendations, even if the nominal probability of matching preferences is equal. This advantage likely stems from access to comprehensive user data, which enhances personalization. Managers of e-commerce platforms like Amazon should capitalize on mandatory collection in markets with low privacy concerns,

using detailed user data (e.g., purchase and browsing history) to refine recommendation algorithms and achieve high ρ^M. To mitigate potential user resistance, managers should implement transparent data usage policies and emphasize the value of personalized recommendations, thereby sustaining user engagement and maximizing ad or sales revenue.

3. **Enhancing Data-Free Recommendations for Competitive Advantage**: The result that profit in Scenario N is higher when $\rho^N > \rho^M$ highlights the potential for data-free scenarios to outperform mandatory collection when generalized recommendations are more effective at matching user preferences. This scenario is particularly relevant in privacy-sensitive markets where users are reluctant to share data. Managers of social media platforms like TikTok can achieve a high ρ^N by leveraging contextual signals (e.g., trending hashtags, geolocation, or device type) to deliver relevant content without relying on user data. Strategies such as A/B testing of recommendation algorithms or incorporating real-time feedback loops can further enhance ρ^N, enabling platforms to outperform Scenario M in terms of profit while aligning with user privacy preferences.

4. **Adapting to recommendation matching degree**: The dependence of profit on ρ^N and ρ^M underscores the need to tailor recommendation strategies to privacy and regulatory constraints, which can influence the feasibility of achieving high ρ. In regions with stringent data protection regulations (e.g., GDPR), mandatory collection may face barriers that reduce ρ^M due to limited data access or user opt-outs. In such cases, managers should pivot to Scenario N, investing in algorithms that maximize ρ^N through non-

personalized data sources, such as market trends or anonymized behavioral analytics. For example, a health management platform could use population health statistics to recommend wellness programs, achieving a high ρ^N while complying with privacy laws. This adaptive approach ensures profitability while minimizing legal and reputational risks.

4.4.2 Consumer Surplus

We categorize three cases:

- Case (1): $\rho^N = \rho^M$.

  Firstly, if $\rho^k \in [\frac{1}{2}, 1]$, then we have:

$$CS^{N*} = \begin{cases} 0, & \text{if } s \in [\bar{s}_2^N, \frac{1}{2}]; \\ s(\rho^N + 1), & \text{if } s \in [0, \bar{s}_2^N] \end{cases}$$

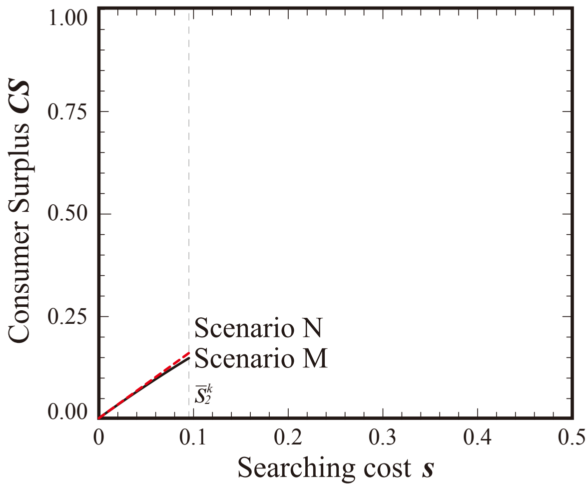
$$CS^{M*} = \begin{cases} 0, & \text{if } s \in [\bar{s}_2^M, \frac{1}{2}]; \\ s(\rho^M - 2\rho^M s + 1), & \text{if } s \in [0, \bar{s}_2^M] \end{cases}$$

  Secondly, if $\rho^k \in [0, \frac{1}{2}]$, then we have:

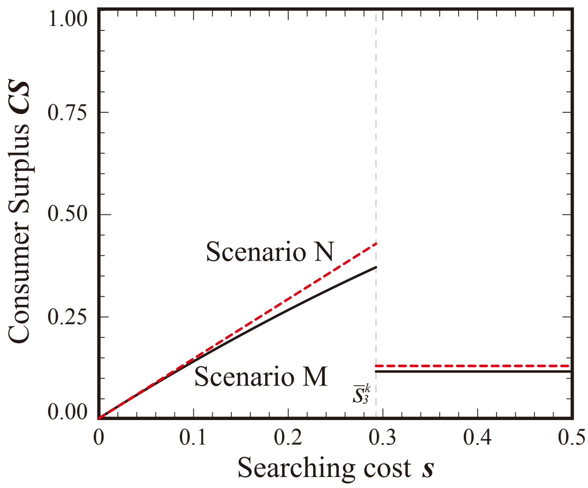
$$CS^{N*} = \begin{cases} \frac{1-4(\rho^N)^2}{8-8\rho^N}, & \text{if } s \in [\bar{s}_3^N, \frac{1}{2}]; \\ s(\rho^N + 1), & \text{if } s \in [\bar{s}_1^N, \bar{s}_3^N] \\ s(\rho^N + 1) & \text{if } s \in [0, \bar{s}_1^N] \end{cases}$$

$$CS^{M*} = \begin{cases} \frac{1-2\rho^M}{8(\rho^M-1)^2}, & \text{if } s \in [\bar{s}_3^M, \frac{1}{2}]; \\ s(\rho^M - 2\rho^M s + 1), & \text{if } s \in [\bar{s}_1, \bar{s}_3^M] \\ s(\rho^M - 2\rho^M s + 1), & \text{if } s \in [0, \bar{s}_1^M] \end{cases}$$

To conclude, by comparison, one can easily verify that $CS^{M*} \leqslant CS^{N*}$ holds when $\rho^N = \rho^M$.



(a) $\rho^k \in [\frac{1}{2}, 1]$



(b) $\rho^k \in [0, \frac{1}{2})$

Figure 4.6: Profit comparison of CS in case of $\rho^N = \rho^M$

- Case (2): $\rho^N < \rho^M$. It follows that $0 < \bar{s}_2^M < \bar{s}_2^N < \frac{1}{2}$ and $0 < \bar{s}_3^M < \bar{s}_3^N < \frac{1}{2}$.

Firstly, if $\rho^k \in [\frac{1}{2}, 1]$, then we have:

$$\begin{cases} CS^{M*} > CS^{N*}, & \text{if } s \in \left[0, \min\{\frac{\rho^M - \rho^N}{2\rho^M}, \bar{s}_2^M\}\right]; \\ CS^{M*} \leqslant CS^{N*}, & \text{otherwise} \end{cases}$$

Secondly, if $\rho^k \in [0, \frac{1}{2}]$, then we have:

$$\begin{cases} CS^{M*} > CS^{N*}, & \text{if } s \in \left[0, \min\{\frac{\rho^M - \rho^N}{2\rho^M}, \bar{s}_3^M\}\right]; \\ CS^{M*} \leqslant CS^{N*}, & \text{otherwise} \end{cases}$$

- Case (3): $\rho^N > \rho^M$. It follows that $0 < \bar{s}_2^N < \bar{s}_2^M < \frac{1}{2}$ and $0 < \bar{s}_3^N < \bar{s}_3^M < \frac{1}{2}$.

  Firstly, if $\rho^k \in [\frac{1}{2}, 1]$, then we have:

$$\begin{cases} CS^{N*} > CS^{M*}, & \text{if } s \in \left[0, \min\{\frac{\rho^M - \rho^N}{2\rho^M}, \bar{s}_2^N\}\right]; \\ CS^{N*} \leqslant CS^{M*}, & \text{otherwise} \end{cases}$$

  Secondly, if $\rho^k \in [0, \frac{1}{2}]$, then we have:

$$\begin{cases} CS^{N*} > CS^{M*}, & \text{if } s \in \left[0, \min\{\frac{\rho^M - \rho^N}{2\rho^M}, \bar{s}_3^N\}\right]; \\ CS^{N*} \leqslant CS^{M*}, & \text{otherwise} \end{cases}$$

1. **Reducing Searching Costs through Effective Recommendation Systems**: The influence of searching cost s on consumer surplus underscores the importance of recommendation systems that minimize user effort in discovering preferred content. In scenarios where consumer surplus is higher (e.g., Scenario $\mathbb{V}$, as suggested by prior results), platforms likely achieve greater recommendation accuracy, reducing s. Managers of platforms like Netflix

should invest in advanced recommendation algorithms that leverage available data—whether comprehensive (Scenario $\mathbb{M}$), voluntary (Scenario $\mathbb{V}$), or contextual (Scenario $\mathbb{N}$)—to deliver relevant content swiftly. For instance, in Scenario N, where user data is limited, platforms can use trending topics or genre-based clustering to lower s, enhancing surplus by enabling users to find desired content with minimal effort.

2. **Leveraging Voluntary Data Collection to Enhance Surplus**: If Scenario $\mathbb{V}$ yields the highest consumer surplus, as implied by earlier results, this suggests that allowing users to voluntarily share data enables personalized recommendations that reduce searching costs more effectively than mandatory or data-free approaches. Managers of social media platforms like TikTok should implement opt-in data-sharing mechanisms, such as prompts that highlight the benefits of personalized feeds (e.g., "Share your interests for a tailored experience"). By fostering user autonomy, platforms can achieve high recommendation accuracy, lowering s and boosting consumer surplus. Transparent data consent interfaces and privacy assurances are critical to encourage participation, ensuring users perceive data sharing as a value-enhancing choice.

3. **Balancing Personalization and Privacy in Mandatory Collection**: In Scenario $\mathbb{M}$, where data collection is mandatory, consumer surplus may be lower if high searching costs persist due to user distrust or irrelevant recommendations. Managers of e-commerce platforms like Amazon, which rely on mandatory data collection, should mitigate this by refining personalization algorithms to align recommendations closely with user preferences,

thereby reducing s. Additionally, implementing privacy-preserving techniques, such as differential privacy, can alleviate user concerns about data usage, indirectly lowering perceived searching costs by fostering trust. By combining precise recommendations with robust privacy measures, managers can enhance consumer surplus in Scenario $\mathbb{M}$, bridging the gap with Scenario $\mathbb{V}$.

4. **Optimizing Generalized Recommendations in Data-Free Scenarios**: In Scenario $\mathbb{N}$, where platforms lack user data, consumer surplus may be constrained by higher searching costs due to less personalized recommendations. To counteract this, managers should focus on contextual and trend-based recommendation strategies that minimize s. For example, a health management platform could recommend wellness programs based on population level health trends or seasonal factors, reducing the effort users expend to find relevant content. Techniques such as collaborative filtering or real-time engagement analytics can further enhance recommendation relevance in Scenario $\mathbb{N}$, increasing consumer surplus by streamlining the user experience despite limited data.

5. **Adapting to Searching Cost Sensitivity Across Markets**: The relationship between consumer surplus and searching cost s suggests that user sensitivity to searching effort varies across markets and demographics. Managers should conduct market research to quantify s (e.g., through user surveys or engagement metrics) and tailor recommendation strategies accordingly. In markets with high sensitivity to s (e.g., time-constrained users), platforms should prioritize scenarios that minimize searching costs, such as Scenario

$\mathbb{V}$ with voluntary data sharing. Conversely, in markets with lower sensitivity, Scenario $\mathbb{N}$'s cost-effective generalized recommendations may suffice. For instance, a streaming platform like Spotify could adjust its recommendation strategy based on regional user behavior, offering highly personalized playlists in high-s markets and genre-based suggestions in low-s markets to optimize surplus.

Chapter 5

Conclusion

Consumer interactions with digital platforms are ubiquitous across online marketplaces. Our study aims to elucidate how different data collection scenarios affect consumer surplus, platform recommendation strategies, and user engagement in a competitive digital environment. Specifically, we focus on three distinct data collection scenarios: the Data-Free Collection Scenario ($\mathbb{N}$), which relies on generalized recommendations without user data, the Mandatory Collection Scenario ($\mathbb{M}$), which enforces comprehensive data access to enable personalized recommendations, and the Voluntary Collection Scenario ($\mathbb{V}$), which allows users to opt-in for data sharing to balance personalization and privacy.

Using a two-stage game-theoretic model, we uncover a nuanced interplay between data collection policies and platform recommendation strategies. Indeed, data policies not only serve to enhance personalization but also significantly influence user polarization and platform service quality. Our results highlight the importance of distinguishing between data-free, mandatory, and voluntary collection scenarios, as each shapes the balance between user privacy and recommendation accuracy differently. Specifically, the level of personalization increases with mandatory data collection but may decrease user surplus in privacy-sensitive contexts, whereas voluntary collection fosters a balanced approach that enhances consumer surplus. Moreover, mandatory collection benefits platforms by enabling precise recommendations, creating a win-win situation when user trust is maintained; however, data-free scenarios can outperform in privacy-conscious markets by leveraging contextual signals, albeit at the cost of reduced personalization. Consequently, platforms' recommendation strategies, pricing decisions, and overall profitability hinge on the prevailing data policy and its alignment with user privacy preferences.

As the primary managerial insights, we link data collection policies with platforms' recommendation strategies and pricing decisions, emphasizing the need to differentiate between data-free, mandatory, and voluntary collection scenarios. Our findings are applicable to various digital industries where data-driven personalization is pivotal and user privacy is a critical concern. Consider the example of social media platforms like TikTok or streaming services like Netflix. In contexts where voluntary data sharing is prevalent, platforms can adopt strategies to encourage user opt-ins by transparently highlighting the benefits of personalized recommendations, such as tailored content feeds or curated playlists. Conversely, in privacy-sensitive markets where data-free approaches dominate, platforms may focus on enhancing contextual recommendation algorithms, leveraging trends or anonymized behavioral data to maintain relevance. For platforms employing mandatory collection, investing in privacy-preserving technologies, such as differential privacy, can sustain user trust, thereby reinforcing the advantages of precise personalization and optimizing both user engagement and profitability.

Furthermore, digital platforms must vigilantly monitor the evolving nature of data collection preferences and adapt their recommendation and privacy strategies accordingly. For instance, in the nascent stages of a platform's lifecycle, such as an emerging e-commerce marketplace, voluntary data collection may dominate as users are more willing to share data to access personalized services, enabling platforms to rapidly refine recommendation algorithms and expand user engagement. As the platform matures and user awareness of privacy concerns grows, data-sharing preferences may shift toward more restrictive behaviors, with users favoring data-free or opt-in models. In this scenario, the emphasis on privacy-preserving recommendations becomes critical, and efforts should focus on leveraging contextual data or anonymized analytics to maintain recommendation relevance. As a TikTok executive noted, "without adapting to users' evolving privacy expectations, platforms risk losing trust and engagement" (TikTok Newsroom, 2022[1]). Thus, platforms should prioritize flexible data strategies that align with shifting user preferences to sustain competitive advantage and capitalize on personalization benefits.

We hypothesize that our main results on the interaction between data policies and recommendation strategies will remain robust, yet the interaction between strategic data-sharing and privacy dynamics may yield novel insights. We leave a comprehensive exploration of this case for future research.

[1] https://newsroom.tiktok.com/en-eu/sharing-an-update-to-our-privacy-policy

References

Ashkan, Azin et al. (2015). "Optimal Greedy Diversity for Recommendation." In: *IJCAI*. Vol. 15, pp. 1742–1748.

Bergemann, Dirk and Alessandro Bonatti (2024). "Data, competition, and digital platforms". In: *American Economic Review* 114.8, pp. 2553–2595.

Chen, Laming, Guoxin Zhang, and Eric Zhou (2018). "Fast greedy map inference for determinantal point process to improve recommendation diversity". In: *Advances in Neural Information Processing Systems* 31.

Choi, W Jason, Kinshuk Jerath, and Miklos Sarvary (2023). "Consumer privacy choices and (un) targeted advertising along the purchase journey". In: *Journal of Marketing Research* 60.5, pp. 889–907.

Cinelli, Matteo et al. (2021). "The echo chamber effect on social media". In: *Proceedings of the National Academy of Sciences* 118.9.

Cookson, J Anthony, Joseph E Engelberg, and William Mullins (2023). "Echo chambers". In: *The Review of Financial Studies* 36.2, pp. 450–500.

Dinerstein, Michael et al. (2018). "Consumer price search and platform design in internet commerce". In: *American Economic Review* 108.7, pp. 1820–1859.

Donkers, Tim and Jürgen Ziegler (2021). “The dual echo chamber: Modeling social media polarization for interventional recommending”. In: *Proceedings of the 15th ACM conference on recommender systems*, pp. 12–22.

Donnelly, Robert, Ayush Kanodia, and Ilya Morozov (2024). “Welfare effects of personalized rankings”. In: *Marketing Science* 43.1, pp. 92–113.

Fainmesser, Itay P, Andrea Galeotti, and Ruslan Momot (2023a). “Consumer profiling via information design”. In: *Available at SSRN 4655468*.

— (2023b). “Digital privacy”. In: *Management Science* 69.6, pp. 3157–3173.

Ge, Yingqiang et al. (2020). “Understanding echo chambers in e-commerce recommender systems”. In: *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pp. 2261–2270.

Huang, Peng, Gaoyan Lyu, and Yi Xu (2022). “Quality regulation on two-sided platforms: Exclusion, subsidization, and first-party applications”. In: *Management Science* 68.6, pp. 4415–4434.

Ichihashi, Shota (2020). “Online privacy and information disclosure by consumers”. In: *American Economic Review* 110.2, pp. 569–595.

Johnson, Garrett A, Scott K Shriver, and Samuel G Goldberg (2023). “Privacy and market concentration: intended and unintended consequences of the GDPR”. In: *Management Science*.

Ke, T Tony and K Sudhir (2023). “Privacy rights and data security: GDPR and personal data markets”. In: *Management Science* 69.8, pp. 4389–4412.

Koh, Byungwan, Srinivasan Raghunathan, and Barrie R Nault (2017). “Is voluntary profiling welfare enhancing?” In: *MIS Quarterly* 41.1, pp. 23–42.

Li, Jin et al. (2023). "Marketplace scalability and strategic use of platform investment". In: *Management science* 69.7, pp. 3958–3975.

Li, Jun and Serguei Netessine (2020). "Higher market thickness reduces matching rate in online platforms: Evidence from a quasiexperiment". In: *Management Science* 66.1, pp. 271–289.

Li, Lusi, Jianqing Chen, and Srinivasan Raghunathan (2018). "Recommender system rethink: Implications for an electronic marketplace with competing manufacturers". In: *Information Systems Research* 29.4, pp. 1003–1023.

Li, Xi, Xin Wang, and Barrie R Nault (2024). "Is personalized pricing profitable when firms can differentiate?" In: *Management Science* 70.7, pp. 4184–4199.

Liu, Dengpan, Sumit Sarkar, and Chelliah Sriskandarajah (2010). "Resource allocation policies for personalization in content delivery sites". In: *Information Systems Research* 21.2, pp. 227–248.

Liu, Jia and Ziwei Cong (2023). "The daily me versus the daily others: how do recommendation algorithms change user interests? Evidence from a knowledge-sharing platform". In: *Journal of Marketing Research* 60.4, pp. 767–791.

Mai, Yunke, Bin Hu, and Saša Pekeč (2023). "Courteous or crude? Managing user conduct to improve on-demand service platform performance". In: *Management Science* 69.2, pp. 996–1016.

Montes, Rodrigo, Wilfried Sand-Zantman, and Tommaso Valletti (2019). "The value of personal information in online markets with endogenous privacy". In: *Management Science* 65.3, pp. 1342–1362.

Pitoura, Evaggelia, Kostas Stefanidis, and Georgia Koutrika (2022). "Fairness in rankings and recommendations: an overview". In: *The VLDB Journal*, pp. 1–28.

Qian, Kun and Sanjay Jain (2024). “Digital content creation: An analysis of the impact of recommendation systems”. In: *Management Science* 70.12, pp. 8668–8684.

Rastegarpanah, Bashir, Krishna P Gummadi, and Mark Crovella (2019). “Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems”. In: *Proceedings of the twelfth ACM international conference on web search and data mining*, pp. 231–239.

Santos, Fernando P, Yphtach Lelkes, and Simon A Levin (2021). “Link recommendation algorithms and dynamics of polarization in online social networks”. In: *Proceedings of the National Academy of Sciences* 118.50, e2102141118.

Shardanand, Upendra and Pattie Maes (1995). “Social information filtering: Algorithms for automating “word of mouth””. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 210–217.

Smedlund, Anssi (2012). “Value cocreation in service platform business models”. In: *Service Science* 4.1, pp. 79–88.

Valletti, Tommaso and Jiahua Wu (2020). “Consumer profiling with data requirements: Structure and policy implications”. In: *Production and Operations Management* 29.2, pp. 309–329.

Wang, Yifan et al. (2023). “A survey on the fairness of recommender systems”. In: *ACM Transactions on Information Systems* 41.3, pp. 1–43.

Wilhelm, Mark et al. (2018). “Practical diversified recommendations on youtube with determinantal point processes”. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 2165–2173.

Xu, Fasheng, Xiaoyu Wang, and Fuqiang Zhang (2025). “Consumer privacy in online retail supply chains”. In: *Management Science*.

Zehlike, Meike, Ke Yang, and Julia Stoyanovich (2022). “Fairness in ranking, part ii: Learning-to-rank and recommender systems”. In: *ACM Computing Surveys* 55.6, pp. 1–41.

Zhong, Zemin (2023). “Platform search design: The roles of precision and price”. In: *Marketing Science* 42.2, pp. 293–313.