University of Milan
University of Pavia

PhD program in: Economics
Cycle: XXXV

Data disclosure and tracking in digital markets: two essays

Supervisor: Prof. Dr. Alberto Cavaliere

PhD Thesis by
Langone Dante
ID number: 479571

Academic Year: 2022-2023

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my parents, Aldo and Miriam, for their unwavering love and support throughout this incredible journey. Their belief in me and their encouragement gave me the opportunity to pursue this achievement. I am forever grateful for their sacrifices and the values they instilled in me.

I am also indebted to my siblings, Alessandro and Giuditta, for their constant love and inexhaustible assistance whenever I needed it. Their presence and encouragement were a source of strength throughout this challenging endeavor.

My heartfelt appreciation goes to my supervisor, Prof. Alberto Cavaliere, whose guidance and unwavering support were invaluable. His dedication to my research, endless hours of discussion, and constructive feedback helped shape my work and push me forward during the toughest times. His guidance, support, and mentorship went beyond the academic realm, creating a nurturing and inspiring environment that allowed me to thrive.

I want to direct special thanks to Prof. Alberto Gaggero for his valuable suggestions, insightful feedback, and the growth opportunities he provided. His belief in my potential and his constant push to strive for excellence were instrumental in helping me "navigate" the challenges and reach the "shore" of completion.

I am grateful to Prof. Tommaso Valletti for engaging in a valuable discussion on the research gap addressed in the second chapter. Their input and insights played a significant role in developing the second article. I deeply appreciate their willingness to share their expertise.

A special mention goes to David, Carlo, Simone, Bob, Tony, Alvaro, Giulia and Federica. Their unwavering support, friendship, and shared moments of joy during these arduous years were invaluable. Your presence made the journey more enjoyable, and I am grateful for the memories we created together.

Finally, I express my sincere thanks to all those who supported me in various ways, both academically and personally, during this remarkable undertaking. Your encouragement, understanding, and belief in my abilities have meant the world to me.

To everyone mentioned here and those whose support and assistance may not be explicitly acknowledged, please accept my heartfelt appreciation for your contributions. This achievement would not have been possible without each and every one of you.

Contents

List of Figures **VII**

List of Tables **IX**

1 Market power and user tracking: an empirical analysis of iOS apps market **7**

1.1 Introduction . . . 7

1.1.1 Research question and main contribution . . . 8

1.1.2 Literature . . . 11

1.2 Data description . . . 14

1.2.1 Data collection: method and variables . . . 14

1.2.2 Apple Privacy Initiative and Privacy indicators . . . 15

1.2.3 Market share proxy . . . 18

1.3 Econometric model and hypotheses . . . 21

1.4 Sample Statistics . . . 24

1.4.1 Summary statistics: data uses . . . 24

1.4.2 Summary statistics: market definition and market shares . . . 25

1.4.3 Controls' summary statistics . . . 27

1.4.4 Summary statistics: Monetization . . . 27

1.4.5 Summary statistics: Updates . . . 28

1.4.6 Other controls . . . 29

1.4.7 Inspection of within variation . . . 29

1.5 Results . . . 30

1.5.1 Main model: the impact of market power on *Data Used to Track You* . . . 30

1.6 Robustness . . . 35

1.6.1 Sensitivity to cluster resolution . . . 35

1.6.2 Other market definitions . . . 35

1.6.3 Endogeneity concerns . . . 36

1.7 Conclusions . . . 38

Bibliography **41**

2 Data Externalities and Vertical Differentiation in Digital Markets: a Welfare Analysis **45**
2.1 Introduction . . . 45
2.1.1 Literature . . . 46
2.2 Basic setup . . . 48
2.2.1 Consumers . . . 48
2.2.2 Firm's profit, stages, and equilibrium concept . . . 50
2.3 First Best . . . 51
2.3.1 Information stage and indifferent consumer . . . 51
2.3.2 Covered market with unaware consumers . . . 54
2.4 Market allocation and welfare with unaware consumers . . . 56
2.4.1 Market allocation and welfare with unaware consumers . . . 56
2.5 Market allocation and welfare with aware consumers . . . 58
2.5.1 Full market coverage case . . . 58
2.5.2 Partial market coverage . . . 61
2.5.3 Comparison covered-uncovered market . . . 64
2.6 Welfare comparison . . . 65
2.7 Conclusions . . . 67

Bibliography **69**

A Appendix A **71**
A.1 Endogeneity and limitations . . . 71
A.1.1 Reverse causality . . . 71
A.1.2 Simultaneity bias . . . 72
A.1.3 Economies of scale and impact of privacy preferences . . . 74
A.1.4 Impact of data on updates . . . 75
A.2 Sensitivity to Market Definition . . . 75
A.2.1 Sensitivity to Resolution Parameter with YMAL network . . . 75
A.2.2 Alternative market definition . . . 76
A.3 Inspection of apps that changed U2TU section . . . 82
A.4 Apps updates behavior . . . 82
A.4.1 Descriptives . . . 84
A.4.2 Results . . . 85
A.5 Regression for data linked to you indicator . . . 87

B Appendix B **89**
B.1 Model with aware consumers . . . 89
B.2 Partial market coverage with aware consumers . . . 91
B.3 First best uncovered market with unaware consumers . . . 94

B.4 First best with aware consumers 108

List of Figures

1.1 Layered structure of privacy information in the Apple App Store 17
1.2 Privacy section in the iOS Store 18
1.3 HHI histogram 25
1.4 Log of *U2TU* by category of market share 27
1.5 In-app purchases indicators for last quartile of L2U and U2TU vs. first three quartiles 28
1.6 Inspection of within variation for main variables 30

2.1 Prices, disclosure and α relationships for different values of $\beta = \{1/10, 9/10\}$ and $\bar{\theta} = \{11/10, 19/10\}$ 61
2.2 Prices, disclosure and α relationships for different values of $\beta = \{1/10, 9/10\}$ and $\bar{\theta} = \{11/10, 19/10\}$ with an uncovered market 63
2.3 Welfare and externality relationships in different models. 66
2.4 Price and α 67
2.5 Disclosure rate and α 68

A.2 Network based on description similarity 79
A.3 Deviation from apps mean *log(U2TU)* vs *months old* 83
A.4 Updates indicators for last quartile of L2U and U2TU vs. first three quartiles 84
A.5 Log of *L2U* by category of market share 87

List of Tables

1.1 Summary Statistics Data Uses 24
1.2 Cluster Level Summary Statistics 26
1.3 Summary Statistics 29
1.4 Selected coefficients for model on *Data Used to Track You* 32

A.1 Sensitivity of the Seller FE model to the resolution parameter and market definition 77
A.2 Sensitivity to market definition based on text analysis 80
A.3 Summary Statistics Data Uses 82
A.4 Selected coefficients for model on update frequency and count of updates 86
A.5 Regression analysis with the log. of data linked to you as dependent variable 88

B.1 Simulation results for monopolist equilibrium with aware consumers . . 91
B.2 Simulation results for planner solution with aware consumers 95
B.3 Simulation results for monopolist equilibrium with aware consumers . . 111

Introduction

Screens have become a pervasive part of our daily lives, with smartphones alone averaging over 4 hours per day. We bounce from screen to screen, including smart TVs, computers, smartwatches, and smart bracelets. These interconnected devices can collect a vast amount of data on our consumer choices and interests. Thanks to advanced Machine Learning (ML) and classification models, firms can use this data to anticipate our searches and infer our interests. While this has reduced search times, improved match quality, and increased innovation, it has also raised privacy concerns.

Over the past decade, there has been a growing interest in the topic of privacy in digital markets. Books such as "Permanent Record" and "The Capitalism of Surveillance" have shed light on the negative effects of data collection. In "Permanent Record," Snowden exposes how governmental bodies collect data, highlighting the issue of digital footprint and government intrusion (Snowden, 2019). Meanwhile, Zuboff delves into the private and business aspects of this world, exposing how powerful companies commodify personal data and manipulate people into revealing more information, all for the sake of profit (Zuboff, 2019).

Only recently, competition authorities have introduced user data and users' privacy concerns within policy cases. In 2019 the German Federal Cartel Office filed a lawsuit and condemned Facebook for the exploitative practices that concerned the collection and use of data from third-party websites and apps that use Facebook's advertising and analytics tools.[1] On the other side of the Atlantic instead, the US Federal Trade Commission filed a lawsuit in 2020 against the abuse of monopoly power in the social media market that originates in Facebook's strategic acquisitions of Instagram and WhatsApp, and the practice of sharing data with Facebook.[2]

During this turmoil, the European Union did not bide its time. It opened the consultations for the Digital Market Act (DMA) and Digital Service Act (DSA) to regulate digital platforms and online services. The DMA designates large online platforms as "gatekeepers" based on market share, size, and impact on competition and provides a new set of rules and obligations to stimulate competition in digital markets.

[1] German Court Upholds Ruling Against Facebook's Data Collection. The New York Times (nytimes.com)

[2] FTC Sues Facebook for Illegal Monopolization. Federal Trade Commission (ftc.gov)

Complementarily, the DSA targets online intermediaries and will assign responsibility for the content provided within its boundaries. It will be crucial for social media companies.[3] Therefore, these recent developments clearly show the importance and increasing attention that the theme of data in digital markets is receiving.

Previous literature has investigated the theme of privacy since the 1980s, producing insights on the uses of consumers' information, the welfare effects of data trading, the biases that consumers face when dealing with personal data, and the effect of privacy regulation on marketing, and more generally on producer surplus. However, due to the invisibility of the phenomenon of data collection and data use to the researcher's eye and due to the increase in world digitization, some themes still need to be completely understood, and investigation is required to intervene with further regulation.[4]

One interesting phenomenon that still needs to be adequately explained is the one of data markups. While this concept was relevant in the Bundeskartellamt Facebook decision, only a few papers try to investigate the relationship between the intensity of data collection and firms' market power. In Chapter 1, I analyze this relationship through an empirical analysis of the iOS App Store, finding a positive association between market shares and data used to track individuals. This finding is interesting because it implies that even in zero-price markets, influential firms find a way to exploit their market power and that despite the presence of privacy notices, due to network effects and market monopolization, consumers may find themselves stuck in low privacy solutions. This motivates the concern and attention of authorities to the theme of consumer exploitation by dominant firms.

In order to investigate this relationship, the first Chapter's research question required the assembly of a unique panel dataset covering 12 months and about 1.2 mln apps available in the Apple Catalog.[5][6] The dataset includes information about apps' descriptive characteristics, proxies for downloads, perceived quality, similar apps, innovation activity, and, most importantly, data uses. On this last element, the Apple Privacy Nutrition Labels provide an impressive granularity of information about the firms' data collection and use practices: they distinguish among different link statuses (linked to the user profile, not linked to it, or used to track consumers) and only for the first two link statuses they provide six alternative data uses (third-party ads, developer ads, product personalization, analytics, app functionality, other purposes) each with the complete set of items collected and exploited (32 categories see Figure 1.1 for a detailed view). Unlike previous literature that focuses on the number of permissions

[3]Digital Markets Act and Digital Services Act: two new EU regulations to address digital challenges | European Commission - European Commission (europa.eu)

[4]The reader can find specific literature reviews within the chapters.

[5]https://apps.apple.com/us/genre/ios/id36

[6]The scraping of data is still ongoing, and the analysis will be replicated on a more extended observation period.

required by an app, this study can proxy data intensity by the purpose of data collection and focuses on the "*Data Used to Track You*" indicator. This different perspective has the advantage of potentially isolating data collection purposes that are most likely to cause an increase in market share from those that aim at extracting consumers' surplus, thus providing a way to break out of feedback loops that have hindered previous studies on this topic.

Innovative techniques were used to establish important explanatory variables, including market shares and concentration indexes. Two methods were utilized: modularity maximization for community identification in a network and text data analysis. A community identification algorithm was applied to the network of suggestions called "You Might Also Like" provided by Apple as a benchmark to determine sub-markets. In order to ensure reliable results, a document similarity matrix was created based on the text analysis of the descriptions, which was then treated as a weighted network. By using modularity maximization, communities were identified. In digital markets where prices are often zero, more than traditional tests like SSNIP is needed, making market definition challenging. Thus, using textual descriptions to capture the feature space of individual offerings makes it possible to define distance among firms by the intersections of their feature spaces. Then, the similarity matrix from text analysis was treated as a weighted network to overcome the inflexible structure of traditional market definition in digital markets. Community identification algorithms commonly used in social and biological networks were then employed. This approach yielded interesting results in the Apple App Store data, detailed in Appendix A.2.2.

Due to the short time dimension and the low within-variation of the panel, the main specification is a Pooled OLS model with categories and sellers dummies to capture time-invariant factors that would influence data use. The former set of dummies is essential to capture the variance stemming from the app's sector: a calculator may need fewer data from external sources than a social network or a search engine to provide the service. The latter, instead, captures the developer's ability and data strategies that may be the same across multiple apps. In the main specifications, although the estimated coefficient for the Herfindahl-Hirschman Index (HHI) is not significant, or its effect on data markup is marginally irrelevant, market share proxies have a significant and positive effect. In the main specification, an increase in market share by 1% is associated with about 0.4% more data items used to track consumers. In the categorized market share specification, dominant and quasi-dominant apps use about 33% more items to track consumers. However, the low within-variation of the panel does not allow for enough statistical power to keep the significance across all of the market share's categories when testing the panel within estimator, and repeating the analysis on a more extended observation period may be needed in future research.

On the other hand, the employment of the within estimator revealed that as apps

age, the amount of data used to track consumers increases. This could be interpreted either as a market power effect or as an app's survival effect. Further research on the relationship between data and entry and exit patterns is necessary to understand whether the increase in data intensity represents a shift in business model or whether there are selection effects at stake motivated by anti-competitive or product improvement uses of data.

In addition to confirming the market power-driven data markup on a new dataset, this Chapter provides descriptive elements about the correlation of data used to track with in-app purchases, app maturity, and to some extent, updates. Apps that track more have a higher number of in-app purchase options and higher average options' prices. This descriptive evidence suggests that apps may provide a set of different qualities in the market to operate price discrimination through in-app purchases by integrating data from different sources and sharing them with third parties. On this theme, I highlight interesting future venues of research in this field that could be analyzed with the same dataset.

On top of the market power effect investigated in Chapter 1, a very recent research strand has shown that influential firms were able to build particularly rich datasets such that ML models can now infer private consumer information by exploiting the correlation among users' types. Since they now only need minimal data about the consumer to infer the remaining fraction, these techniques allow firms to exploit consumer data despite privacy regulations. Thus, in cases where consumers have heterogeneous privacy preferences, the effects of this negative externality on welfare remain to be understood entirely.

In Chapter 2, I propose a theoretical model to explain how the correlation among consumer data impacts social welfare when a monopolist faces consumers with heterogeneous privacy concerns. The Chapter presents a model where data is a quality element of a monopolist platform with two revenue instruments: prices and data disclosure to advertisers. The negative externality is added as a network effect proportional to the stock of information accumulated by the firm and inversely proportional to the level of privacy the platform offers. With this approach, I adapt the traditional theme of the optimal quality decision of a monopolist that faces heterogeneous consumers to the novel framework provided by the economics of privacy literature, where consumer data constitutes both a revenue source alternative to prices and an element of quality.

This model further reinforces the findings of the previous Chapter and shows that even when the price is zero, there is a welfare loss associated with the monopoly that stems from an under-provision of privacy. This is relevant because digital platforms have commonly used the zero-price argument to defend themselves in court and to show that consumers could not be exploited. Instead, this result further motivates the decision of the Bundeskartellamt against Facebook, which is itself a zero-price

platform. Furthermore, this analysis highlights the inappropriateness of the SSNIP test because, from the model, it arises that the firm may set a zero or negative price (subsidizing consumers), and it exploits its market power through a higher disclosure rate. Therefore, in such cases, the SSNIP test would not capture the profitability of a price increase, and a Small but Significant and Non-transitory Decrease in Quality (SSNDQ) would be a more appropriate tool.

Another set of results from the model is obtained by studying the impact of the externality on welfare. The second essay analyzes two cases: in the first, consumers are unaware of the externality, while in the second, they can perfectly anticipate its damaging effects and consider them in the joining decision. Generally, as expected, the negative externality is welfare detrimental with respect to the no externality case.

However, somewhat unexpectedly, adding the negative externality produces a welfare increase to the no externality case when data correlation among users is strong enough, consumers can anticipate it, and they have a viable outside option. In this (limited) case, the disutility from the data correlation contributes to reducing the quality distortion of the monopolist - also known in the literature as Spence Distortion - by increasing the demand elasticity to data disclosure above the demand elasticity to prices that pushes the monopolist to switch to the price instrument optimally. Eventually, for large enough externality values, the consumers' average willingness to pay for quality becomes equal to the marginal willingness to pay, the Spence distortion disappears, and the only welfare loss that remains stems from the price markup.

This second result has implications for the choice of the optimal policy. Although the proposed strategy of decorrelating data would be an effective solution against the loss that derives from the externality, it still needs to repair the under-provision of quality. When the externality has a strong impact and consumers have an outside option to avoid it, the model shows that raising awareness about the impact of others' privacy decisions may raise the demand's sensitivity to privacy enough to push the platform to offer a higher quality alternative.

However, this Chapter suffers some critical limitations that shall be expanded by future research. The policy debate currently revolves around how to regulate privacy and data markets. While the study has implications on the effectiveness of decorrelating data, it is generally silent about regulation. The model could then be used in future research to study the effect on the welfare of alternative data regulations, such as a cap on the disclosure rate (analogous to a minimum quality standard), a Pigouvian tax on disclosure revenues, or data decorrelation/anonymization. Secondly, the descriptive evidence of the empirical analysis suggests that there may be a firm quality differentiation. The firm offers premium and privacy-preserving features that consumers with different willingness to pay for privacy may be able to buy. In that case, the model result shall be adapted to multiple offerings by the same firm to study

the effect of the externality on welfare. Finally, the location-then-price model with duopolists quickly becomes algebraically irksome. Therefore, the effect of competition remains unexplored in the presence of the informational externality. A way out that could be considered in future research would be to constrain the price of the service to zero and assume that firms only produce revenues from advertising. This would limit the model's generality but allow studying the impact of competition when the externality is present.

Chapter 1

Market power and user tracking: an empirical analysis of iOS apps market

1.1 Introduction

The pervasiveness of smartphones in consumers' life has increased drastically in the last ten years. Smartphones are now the primary data collection device, and some websites report the fact that globally, people average 7 hours of screen time per day, and more than half is represented by smartphones.[1]
Consequently, consumers access apps for any task: to run business meetings and lectures, they use Zoom calls and digital whiteboards, to meet someone, they resort to social networks and dating apps, their television and entertainment are now app-based and provided by Big Tech companies, and they even have an app for managing their pets' dating life.[2] Thanks to this increase in consumption, the market size of the apps' global market has already reached US$430.90bn in 2022 (size comparable to the EU sales in the 'Passenger Car Market'), and it is projected to reach a market volume of US$641.10bn by 2027 depicting a truly astonishing success of the digital economy. Surprisingly, however, most apps offer a zero-priced base product, and developers have found other ways to generate such significant market revenues: selling advertisement slots, trading consumer data, and selling in-app purchases. Therefore, data has assumed both the role of currency and input in the production process and allows consumers to access free services (Kummer and Schulte, 2019; Cecere, Le Guel, and Lefrere, 2020).

With data assuming the role of currency, the concern that larger firms impose a higher data markup and exploit consumers was raised in courts. Due to both consumer biases and to the absence of competition, the attention of competition authorities

[1] https://www.bbc.com/news/technology-59952557. Other source CDC infographics: link
[2] Here is the page of one of the many presents on the Apple App Store: link

shifted to privacy terms. [3]

Indeed, empirical research in the economics of privacy has shown that consumers suffer many behavioral biases when trading personal data to access free services, and despite the General Data Protection Regulation (GDPR), influential firms have been able to use dark patterns to extract data from consumers (Acquisti, Brandimarte, and Loewenstein, 2015; Norwegian Consumer Council, 2018). Nonetheless, the empirical literature on the presence of data markups is scarce due to the difficulty in finding adequate proxies for market power and data strategies and the presence of endogenous feedback loops.

This paper investigates the relationship between market power and data markups by studying a novel panel dataset assembled from the Apple App Store. The paper contributes by proposing an analysis focusing only on data uses that are most likely focused on surplus extraction and less affected by the concern for reverse causality. Furthermore, the paper provides a methodological contribution to the definition of digital markets by creating competition proxies through network science and text analysis. The paper's results confirm the ones found in the previous literature, and it shows that the concern for consumer data exploitation is well founded, and market power is associated with higher data markups even after controlling for developers' fixed characteristics and apps-specific controls.

The structure of the paper is as follows: subsection 1.1.2 discusses the related literature and subsection 1.1.1 better defines the research question, Section 1.2 presents the dataset and some descriptive statistics. Then, Section 1.3 illustrates the econometric model employed, and Section 1.4 highlights in-depth sample statistics and descriptive evidence for the data used. Finally, Sections 1.5 and 1.7 conclude with the results and their interpretation. Complementarily, Section 1.6 provides the limitations and the assumptions needed and reports the results of some sensitivity tests, whose more extensive treatment can be found in Appendix A.1. Among these are the robustness of market definition and a small extension that analyses the impact of different data uses on the updating process.

Before diving deep into the analysis, the research question and related literature are presented in the following two subsections.

1.1.1 Research question and main contribution

Analyzing the relationship between market power and data in the digital economy is complicated by several factors. These include the challenge of defining markets in zero-priced digital markets, the low observability of firms' data strategies, and the

[3] German Court Upholds Ruling Against Facebook's Data Collection. The New York Times (nytimes.com)

presence of endogenous relationships and feedback loops between data and market structure.
However, the availability of text data, faster computing power, and recent attention to privacy regulation contribute to reducing these difficulties in three ways. Firstly, the introduction of the General Data Protection Regulation (GDPR) increased market transparency about data strategies by introducing the obligation to state the purpose of data collection within the privacy notices proposed to consumers. Secondly, as discussed in section 1.2.2, the Apple App Store, with its recent Privacy Nutrition labels, forces apps to specify data use in a schematic and salient way that potentially helps to avoid feedback loops. Thirdly, advancements in the realm of networks and communities identification supported by an exponential increase in computing power gave researchers the ability to explore new methodologies to proxy competition, such as modularity maximization in network analysis (Newman, 2006) and Natural Language Processing that a decade ago were not computationally tractable for large datasets like the one used in this article.

With the increase in exploitation and remuneration of consumer data in secondary markets, the EU regulator has intervened with the GDPR, which has been taken as a model worldwide. This regulation was needed because of the many biases consumers suffer. However, it was not enough to avoid the fall of consumers into dark patterns and the resulting lack of control over personal data, despite the many privacy notices they are required to sign when accessing any service online (European Commission, 2019; Acquisti, Brandimarte, and Loewenstein, 2015; Norwegian Consumer Council, 2018).

Therefore, the widespread discussion about intervention versus the laissez-faire approaches comes back also in this field. Can competition in privacy terms push firms to offer better privacy terms where regulation fails? Moreover, do more powerful firms use data more intensively due to their market power? To answer these questions, we must understand how different data uses are affected by competition and how they impact competition. This article describes the relationship between market power and the amount of information used to target consumers.

This research establishes a correlation between market power and the use of data by different companies, which confirms what other studies have also found (Dimakopoulos and Sudaric, 2018; Kesler, Kummer, and Schulte, 2019; Preibusch, Kübler, and Beresford, 2013). The approach and methodology align with other researchers (Kesler, Kummer, and Schulte, 2019), but are applied to a novel panel dataset obtained from the Apple App Store market. The article's results show that companies with larger market shares tend to track consumer data more extensively, and this trend is consistent across various market definitions and functional forms.

While the previous literature analyzing app privacy focuses on apps' permissions,

this article exploits the unique information from the Apple App Store that splits among different data uses and "link statuses". This difference lets us focus on data uses with a higher surplus extraction component. So under the hypothesis of non-simultaneity in the Privacy Label's choice, the typical reverse causality problem would be attenuated. In fact, the characterization of different data uses in the privacy nutrition labels allows us to disentangle the cases where data has a surplus extraction term that balances the markup effect in the terminology of De Cornière and G. Taylor (2023).[4] In the Apple environment, we can distinguish between data uses that are unilaterally pro-competitive (UPC) and potentially reverse causal (product personalization, app functionality, analytics, and potentially third-party advertising) from cases where data use would not be endogenous (tracking). Furthermore, privacy preferences can be easily influenced by the way information is presented, as different framing can elicit varying levels of concern. Studies have shown this to be true (Tsai et al., 2011; Acquisti, Brandimarte, and Loewenstein, 2015). In turn, consumers' privacy concerns may be activated by the label *Data Used to Track You*, as better defined in Section 1.2.2, which is the most worrisome for privacy. Consumers' privacy preferences, added to the surplus extraction component, make this data use less likely to be *UPC* and a clear expression of market power.

Additionally, the sample descriptives and the regression analysis show that the apps that use more data to track consumers across apps are associated with a higher number of in-app-purchases and a higher mean value of in-app-purchases (see Section 1.4). Therefore, apps may be using these data items to tune their pricing options, and this may involve some elements of price discrimination among different demand elasticities and multiple qualities: that would represent a data use with high surplus extraction term, so also on this front, a lower probability of satisfying the conditions stated in De Cornière and G. Taylor (2023). [5]

Finally, I contribute methodologically with an innovative way to cluster apps: I exploit recent advancements in computing power, network science, and text analysis to provide robustness to the market definition employed in Kesler, Kummer, and Schulte (2019). To define markets, I employ the network analysis through modularity maximization of the 'similar apps network' (You Might Also Like - YMAL) proposed in Kesler, Kummer, and Schulte (2019), providing sensitivity to some parameters in

[4] A thorough revision and application to this context of De Cornière and G. Taylor (2023) is presented in the appendix A.1.1

[5] Concerning data used for third party advertising De Cornière and G. Taylor (2023) shows that there are multiple conditions for this data use to be Unilaterally Anti Competitive (UAC). Among them, the assumption that the firm does not use both the price and the advertising level may not hold in the app market, where despite prices being constrained to zero, firms can use in-app purchases to tune the monetary instrument and the advertising one. Therefore, I only consider *Data Used to Track You* as an exogenous indicator because privacy concerns make it less likely to be UPC. We further discuss the implications and necessary conditions not to have endogeneity in estimating the market power effects on data uses in section A.1.

the algorithm as discussed in Sections 1.2.3 and 1.6. Interestingly, Apple does not provide the 'You Might Also Like' section for all their native apps (iMessage, Apple Podcasts, Apple Books, ...). Additionally, given that this information is mostly based on the download patterns of users, the entrants may not be classified optimally. So as sensitivity analysis, I propose an alternative way to define markets using text data, and I expand a recent strand of literature that proxies for competition by using text data. This approach is drawn from the literature that aims at measuring market competition through the intersection of the feature spaces of firms' offerings (Hoberg, Phillips, and Prabhala, 2014; Pellegrino, 2023; Hoberg and Phillips, 2010). Recently, Leyden (2018) defined categorical markets by analyzing product descriptions and clustering through unsupervised machine learning techniques. I employ a similar methodology that this literature has used to extract a similarity matrix that captures the distance of firms in the feature space, as further detailed in 1.2.3 and Appendix A.2.2.

1.1.2 Literature

The economics of mobile apps literature covered the estimation of apps' demand and the factors influencing it. For example, Ghose and Han (2014) estimates a positive correlation of demand with the in-app purchase option, app age, and number of apps of the same developer, among other factors. Furthermore, Kummer and Schulte (2019) expanded the subject by studying the impact of privacy permissions on demand and highlighted the role of the preferences for privacy in the market, finding that there exists substitutability between apps' privacy-intrusiveness and their price. Additionally, Bian, Ma, and Tang (2021) showed that when apps' privacy information is published, apps with more invasive data strategies have a higher drop in downloads.

Instead, a cluster of recent articles investigates the updates of mobile applications. Yin, Davis, and Muzyrya (2014) shows the differences between successful game developers and non-game developers. They highlight that while the former category has a higher chance of success with sequential innovation by producing more apps and not working on updates, incremental innovation through frequent updates raises non-gaming apps' likelihood of success. Instead, Comino, Manenti, and Mariuzzo (2019) focuses on the strategic use of updates to increase downloads and shows how successful developers in the Apple App Store tend to use updates to counteract a slowdown in downloads. Finally, Leyden (2018) classifies updates into bug fixing and feature addition through text analysis, and it uses this information to provide a structural model of product updating decisions and developers' innovation in the productivity apps category. Although the main focus of the present analysis is on market power and data, due to the impact of data on updates and consequentially on competition, I contribute to this literature by providing descriptive shreds of evidence of the relationship

between data uses and update behavior.

Another strand of literature in the economics of apps has focused on the impact of data on the choice of business models. A peculiarity of the digital economy is the availability of apparently free services and remaining profitable at zero price required to find new ways to increase revenues. Apps in these markets pick one out of three types of business models: paid, "freemium" with in-app purchases, and free with data trades and ad-funded (Cecere, Le Guel, and Lefrere, 2020). As with respect to worldwide app revenues, in 2022, 51% of the global turnover was earned through advertising and data sales, while 47% was earned through in-app purchases and only a slim fraction derived from the upfront apps price.[6] Therefore, consumer data has assumed the role of currency, an idea that this literature has taken into consideration both theoretically and empirically.

Casadesus-Masanell and Hervas-Drane (2015), which characterizes data/privacy features of online services as quality differentiation elements takes the substitutability of prices and data as potential sources of revenues theoretically into account. This paper explains duopolists' (revenue) differentiation decisions and relates them to the willingness to pay for the product and the heterogeneity of consumers' privacy preferences. The theoretical model shows that while a monopolist exploits both streams of revenues, when competition increases, each firm decides to differentiate and specialize on only one revenue source (data or prices). In this case the substitutability between prices and data disclosure appears.

Empirically, Kummer and Schulte (2019) was the first to estimate the 'data-for-money trade-off' using Play Store app's observational data. This paper focuses on the 'data as currency' by confirming the impact of sensitive permissions (grade of privacy intrusiveness of an app) in substituting prices. Furthermore, it identifies a lower bound in the reduction in demand for an app that requests more privacy-sensitive permissions, showing that consumers have positive privacy preferences. Related to this topic, Bian, Ma, and Tang (2021) estimates the impact on the app's download of the recent Apple Privacy Nutrition Labels initiative. By employing a difference-in-difference approach, this working paper finds a significant reduction in downloads after introducing the privacy summary proportional to the data collection intensity. Similarly, I exploit the information in the privacy labels better described in Section 1.2.2.

This last cluster of research falls in the intersection between the economics of mobile apps literature and the economics of data and privacy literature that instead has investigated the use and the profitability of data collection theoretically. The early economics of privacy has studied the effects of data collection on firms' behavior from various perspectives, highlighting a richness of potential data uses (Acquisti, Brandimarte, and Loewenstein, 2015). Information can be exploited by firms for surplus

[6]Statista website

extraction, such as price discrimination (Fudenberg and Tirole, 2000; C. R. Taylor, 2004; Acquisti and Varian, 2005; Calzolari and Pavan, 2006), to increase demand and shift revenues upwards through product improvement and personalization (Acquisti and Varian, 2005), to improve revenues coming from advertisements through better targeting (De Cornière and De Nijs, 2016). Finally, with the growth of data intermediaries and aggregators, information can be sold in secondary markets (Montes, Sand-Zantman, and Valletti, 2019). [7]

In the digital world, data plays a similar role to prices in traditional markets. This raises the question of whether dominant companies can increase their profits by collecting and utilizing more data than their competitors. This question resembles the first correlational studies in empirical industrial organization that aimed at estimating the link between market power and prices (Schmalensee et al., 1989). However, while the interest in the functioning of digital markets is at its peak, few observational studies assess the strength of the correlation and the causal link between market structure, market power, and user data exploitation by influential firms.

A recent strand of articles investigates the relationship between data and competition. Some articles try to estimate the effects of market power on data markups (Kesler, Kummer, and Schulte, 2019; Preibusch, Kübler, and Beresford, 2013), while others try to highlight the impact of data on the long-term industry dynamics (Prüfer and Schottmüller, 2021; Farboodi et al., 2019). This potential feedback loop requires attention in empirical studies due to the problem of reverse causality.

The theoretical link that may drive firms that face lower competition to have higher data markups has been investigated in Dimakopoulos and Sudaric (2018), which shows how softer competition on either side of the market leads to increased data collection.

This theoretical result and the 'data as currency' phenomenon motivates the study Kesler, Kummer, and Schulte (2019), where the authors investigate the relationship between competition and apps' number of requested permissions in the Google Play Store. With a large sample of more than two mln observations followed quarterly over a two-year window, they study conditional correlations in cross-sections and panel regressions that seem to confirm the positive relationship between market power and data markups reported in the literature (Dimakopoulos and Sudaric, 2018; Bian, Ma, and Tang, 2021; Preibusch, Kübler, and Beresford, 2013). Additional to panel regressions, they provide an empirical strategy based on apps' re-categorizations and exploit them as exogenous variations in market power. Therefore, they are able to solve the reverse causality problem and confirm the positive relationship between market power and data extraction found in fixed effects regressions.

[7] In-depth review of this literature is in Bergemann and Bonatti (2019), Bergemann, Bonatti, and Gan (2022), and Acquisti, C. Taylor, and Wagman (2016) and Acquisti, Brandimarte, and Loewenstein (2015)

Preibusch, Kübler, and Beresford (2013) analyzes the privacy policies of 140 web retailers across five industries. While their results on the data for money trade-off are not so unambiguous as the ones in Kummer and Schulte (2019), they find a negative correlation between data intensity and the amount of direct competitors a website has.[8]

However, their paper could not solve the reverse causality problem behind the market structure and data extracted relationship: is the website extracting more data because it has no competition, or has it no competition because it is extracting more data?

On this issue, the recent working paper of De Cornière and G. Taylor (2023) proposes a general model of competition in utility. It shows that data uses with a higher surplus extraction term, or those that elicit more substantial privacy concerns, are less likely to impact market shares and could express market power. The proposed applications of this model distinguish different data uses by the magnitude of the surplus extraction term: for example, data for product personalization would have an impact on next period market concentration because the surplus extraction term is zero (data is *Unilaterally Pro Competitive* - UPC), but other data uses such as data used for price discrimination incentivize the firm to offer lower utility (due to the surplus extraction term) so that data would not affect next period market share.

1.2 Data description

1.2.1 Data collection: method and variables

The data for this paper contains publicly available information that has been scraped from the Apple App Store. The dataset has been collected through a Python crawler from the catalog of Apple App Store. The Python scraper collected all the apps' links then it opened one-by-one each app page to collect information. [9] The full scrape needed ten days for each wave collected, and the extraction order remained similar along all the waves. With this procedure, six unevenly spread waves were collected from 01/12/2021 to 01/12/2022.

[8]Probably the data for money trade-off is not supported because of their selection methodology: they argue that, since individuals display horizontally differentiated preferences across different data items (e.g. some may prefer to share health-related data while others may be averse sharing it and would rather share hobbies or some other data category), information extraction cannot be considered as a vertical differentiation framework and retailers that extract different data items cannot be directly compared. Therefore, they analyzed only those retailers that extract with different intensity the same category of data, and they compared their prices and privacy policies. It turns out that with this sample selection rule, retailers with higher privacy also correspond to those with lower prices.

[9]We observe entrants after some months from their release, and their number does not correspond to the astonishing figure released by professional data collectors (almost 2000 daily new apps). However, we assume that at least the largest apps are listed on the catalog in at least one category.

I have gathered a dataset containing information on 1.2 million apps that were tracked for 12 months. With information about:

- App **identifiers**: unique ID, name of the app, developer name, date of extraction
- App **descriptives**: description, store category, languages, age rating (PEGI), list of updates, date of release, number of ratings, average rating, and rating distribution (e.g., the fraction of rates by rating level), price, in-app purchases, estimated downloads, estimated revenues, release date and vector of similar apps ID
- **Privacy Labels**: as described in the next subsection

Exclusion criterion and sample size

After the creation of the key variables (privacy indicators, developer statistics, and market shares), the apps that satisfied these conditions were dropped from the sample:

i apps that do not have an iPhone or iPad version (only Mac or iPod);

ii apps that do not have at least 20% of their description in English;

iii gaming apps;

iv apps that did not provide the privacy summary in at least a wave;

The final sample is an unbalanced panel with 434955 app ids followed in six waves from 1/12/2021 to 1/12/2022 for a total of about 2.4 mln observation.

1.2.2 Apple Privacy Initiative and Privacy indicators

In the effort to differentiate itself from the Android ecosystem, Apple is pushing towards the title of privacy champion and has introduced with iOS 14.2 a series of means to provide users with the tools and information for not being tracked.[10] It has introduced the Apple Tracking Transparency (ATT) feature, under which apps must now request consent from users to track them across apps, and the Apple Privacy Nutrition labels that the developer must provide to deliver an update of the app. [11]

[10]It is out of the purpose of this paper to discuss whether this is a strategic initiative to leverage their power to increase their ad revenues cutting out other strong competitors such as Google and Facebook, or it is only a way to differentiate in the spirit of Etro (2021).

[11]I assume information is truthful and do not consider the strategic release of information. However, some technical articles report that privacy summaries may not be as truthful as they seem. An informative version is: Apple Privacy Nutrition Labels. Therefore, the effect of strategic use of this information is possible, and section 1.6.3 presents a discussion of the error introduced by the relaxation of this assumption.

The nutrition labels are self-reported summaries of the privacy policies that increase the saliency and observability of the data strategies of firms. The amount of information provided is massive and structured in several layers, as described in Figure 1.1. In the first layer, the user finds four different data types or also called 'link statuses' '*Data Used to Track You*' (hereafter *U2TU*), '*Data Linked to You*' (hereafter *L2U*), '*Data Not Linked to You*' (hereafter *NL2U*), '*Data Not Collected*' plus a fifth *Summary not Provided* (that does not appear in the Figure because, as a preliminary step for the analysis, we discarded those apps that have not provided the privacy summary).

The main distinction between the *U2TU* and *L2U* link statuses is that information collected in the first is also shared among other apps, data brokers, and ad networks providers, while the second can be used for advertising or other purposes internally, without transferring consumer information to third parties. So suppose a developer provides an ad space: he can share personal data to an ad network manager that efficiently allocates the spaces across multiple apps, or he may auction a spot for some characteristics of the consumer without transferring the actual data to a third party. In the first case, the developer must list the item in the *U2TU* section; in the second, he must list the data item only in the *L2U* section. [12] This distinction is the most important one for this article because different data uses have different relationships with competition, and the ability to identify the data uses with such precision is the core identification strategy I follow.

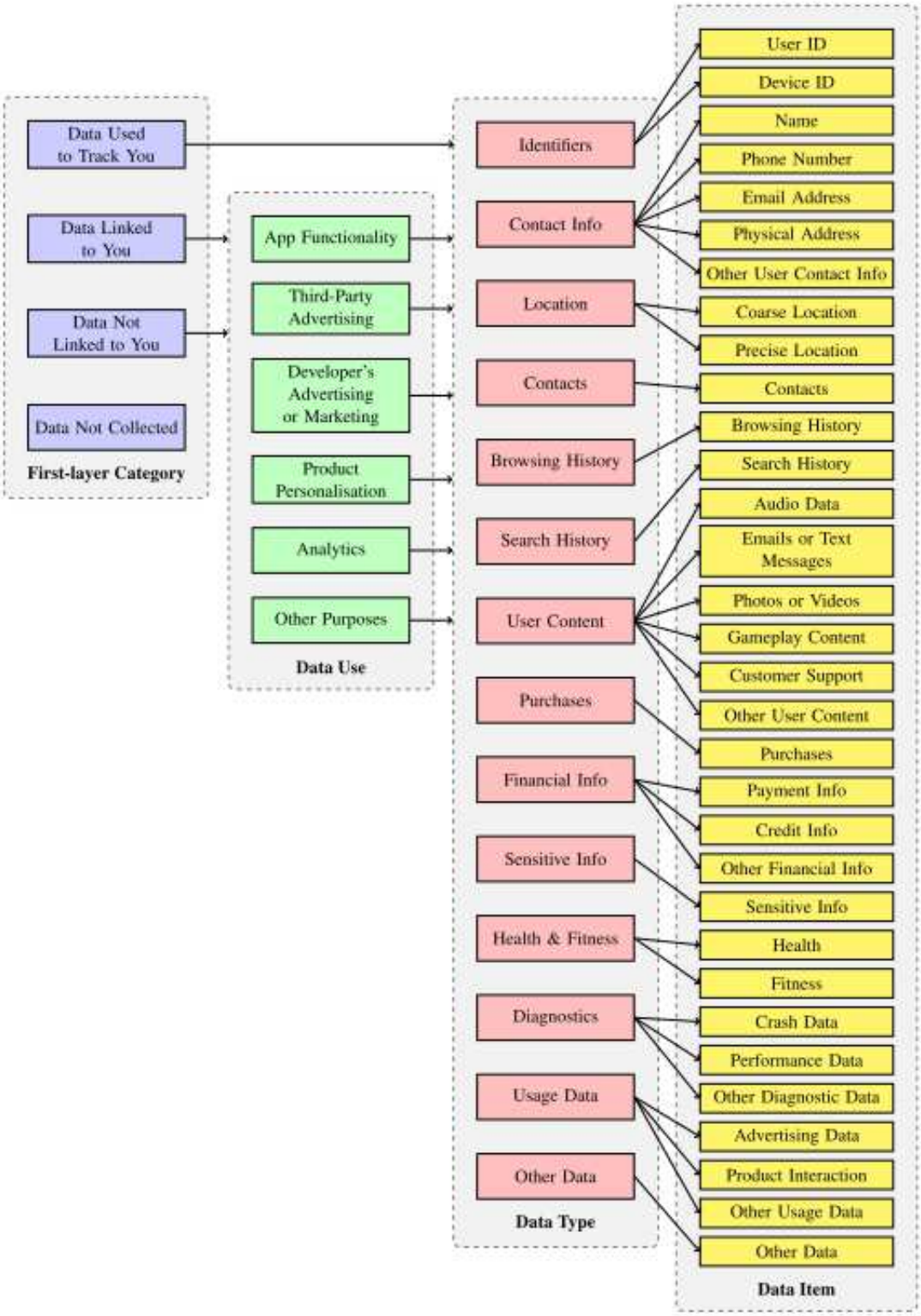
The second layer is available for *L2U* and *NL2U*, describing the purpose of data collection. This is divided in five specific categories: *App Functionality* (*af*), *Third-Party Advertising* (*tpa*), *Developer's Advertising or Marketing* (*da*), *Product Personalization* (*pp*), *Analytics* (*ana*) plus a sixth one being more vague *Other Purposes* (*other*). The third layer displays the 14 possible data types that an app may collect, which in the fourth layer are disaggregated into all possible data items for each data type. This final is the most detailed view of the app behavior with a granular view of the 32 data items.

A view of how this information is presented on the main page of an app in the Apple App Store accessed from an iOS smartphone is depicted in the left Figure 1.2. In the same Figure the two screenshots on the right show the "see details" and the breakdown of privacy policies.

This information has been scraped and converted into a set of dummies coded as '*status_use_type_item*' that are equal one if that particular combination of link status, use, data type, and data item is present (e.g., *l2u_tpa_identifiers_userid* is equal to one if the user id is linked to the personal profile and it is used for third party advertising).

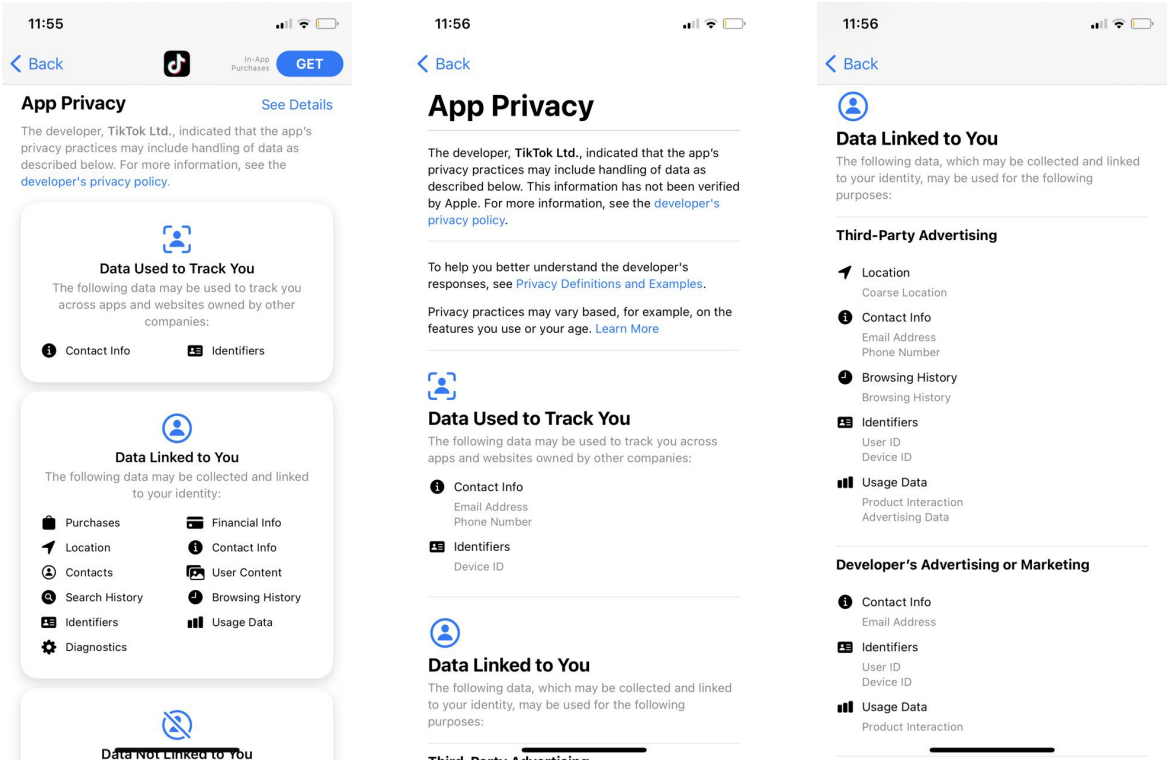
[12] The information here presented is as taken from the Apple guidelines to fill and explain the privacy labels: at Apple's Privacy Labels.

Figure 1.1: Layered structure of privacy information in the Apple App Store



The first-layer indicates the *category*, the second layer the *use*, the third indicates the *type* each composed by different set of items. Source: Bian, Ma, and Tang (2021) Figure 2

Figure 1.2: Privacy section in the iOS Store



The disaggregation returned more than 430 unique dummies that have been aggregated to form privacy scores representing the intensity of the different data statuses. Notice that the app may collect the same data for different purposes, and therefore multiple dummies may stem from the same item.

Data Link status-Use intensity Raw indexes describing the privacy policies have been constructed by summing the dummies: all the dummies that start with a combination of link statuses ($l2u$ or $nl2u$) and data use (af,tpa,da,pp,ana,$other$) have been aggregated to form the correspondent index ($l2u\_af$,$l2u\_ana$,$l2u\_pp$,$l2u\_tpa$,$l2u\_da$,$l2u\_tot$ and $u2tu\_tot$).[13]

The $U2TU$ indicator will form the dependent variable as it is the one that is most likely to be impacted by market power and not to affect it.

1.2.3 Market share proxy

In order to find out which apps are competing with each other, Kesler, Kummer, and Schulte (2019) utilized a community identification algorithm that employed modularity maximization. They did this by using Google suggestions for "Customers also

[13]Given that the min (0) and the max (32) score of these indicators is the same for every one of them it has been taken not standardizing it

bought" to build a network. Similarly, our study utilizes information from Apple's "You Might Also Like" (YMAL) suggestions for similar app IDs. We then treat the similar apps as a network and apply the Blondel et al. (2008) algorithm for modularity maximization, as described in the next paragraph.

It is worth noting that the YMAL network does not suggest similar apps for all of Apple's apps. It could be because Apple strategically promotes its apps and may want to avoid suggesting competitors' apps like those from Google or Microsoft. Whatever the reason, it may create a biased market definition. Therefore, an alternative market definition that does not rely on the YMAL suggestion network is proposed as a robustness test. [14]

We used the modularity maximization algorithm to analyze the YMAL network and identified 522 communities. We also applied the same algorithm to the alternative network created from the apps' descriptions and found 3209 communities. To ensure the clusters were accurate, we manually validated them by analyzing the network of the highest-rated apps. Market definition is a highly contested task in competition economics, and our proposed technique is not exempt from criticism. One drawback is the absence of a clear metric for checking the appropriateness of market definition. To address this, we provide sensitivity analysis to the tunable parameters in Section 1.6 and an alternative market definition in the Appendix. The following paragraph discusses the details of the modularity maximization procedure and the algorithm employed.

Modularity Maximization - Louvain algorithm The modularity maximization technique is frequently employed in network science to obtain the network's community structure. As an optimization-based method, it involves maximizing the so-called 'modularity' of the network defined as

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{1.1}$$

Where m is the total number of links, $A_{i,j}$ the adjacency matrix $k_{i,j}$ are the degree of node i and j and the δ is the Kronecker delta with C_i that represent the class labels of the community to which node i belongs to. This measure compares the number of edges within communities ($\frac{1}{2m} \sum_{i,j} A_{ij} \delta(c_i, c_j)$) to the expected number of edges in an equivalent network (with the same degree distribution) with randomly placed edges ($\frac{1}{2m} \sum_{i,j} \frac{k_i k_j}{2m} \delta(c_i, c_j)$).

If the number of edges within communities is larger than the expected number in an equivalent network with randomly placed edges, then the Q is positive. A large and positive value of Q indicates the possible presence of community structure. Reversing this observation, we can look for community structure by changing the

[14] For more details about this alternative definition, please refer to Section 1.6.2 and Appendix A.

network divisions in sub-communities and selecting the divisions that maximize the value of Q.

Computational algorithms are the quickest method to find a network division with maximum modularity. One such algorithm, described in Blondel et al. (2008), utilizes a bottom-up approach to optimize modularity. The algorithm begins by assigning each node i to a separate community, which is then merged with their j closest neighbors. Next, modularity changes are calculated until a local maximum is reached, and this process is repeated for every node until the global maximum is achieved. The network is then recalculated, and the procedure is iteratively repeated until modularity begins to decrease.

The analysis presented here uses the command *cluster_louvain* in R's package *igraph* to solve this partitioning problem. Given that the algorithm is hierarchical, setting the resolution parameter is possible. This parameter determines the level of detail in community detection within a network, and it controls the granularity of the process, allowing for more specific identification of communities. At a technical level, the resolution parameter affects the optimization process of the algorithm by adjusting the balance between modularity and resolution limit. Modularity seeks to maximize the density of connections within communities while minimizing connections between communities. The resolution limit controls the trade-off between maximizing modularity and allowing for the detection of smaller communities by penalizing the creation of new communities. A higher resolution value leads to more fine-grained communities, while a lower value results in larger, more general communities. By adjusting the resolution parameter, users can explore different scales of community structure within a network, allowing them to analyze the network's organization at varying levels of detail. The resolution parameter has been chosen by validating the clusters by hand. However, the sensitivity of the results to this parameter is reported in Section 1.6.

Market shares To proxy the number of downloads for an app, we can use the number of ratings as a minimum benchmark since users need to download the app to rate it. Research has demonstrated that this measure strongly correlates with the actual number of downloads (Kummer and Schulte, 2019). Hence, in line with previous literature on app markets, we adopt the rating count as a substitute for the number of downloads. This is represented by the formula:

$$s_{i,j,t} = \frac{r_{i,t}}{\sum_{j,t} r},$$

the market share of i in market j is the ratio between the rating and the sum of all the ratings in that cluster. [15] [16]

1.3 Econometric model and hypotheses

In this study, I investigate the impact of market power on the data collected. Given the recent iOS update that forced developers to publish the Privacy Nutrition Labels, we can distinguish between different data uses. I focus on *Data Used to Track You* indicator $U2TU$ because, for this particular data use, the effect of data on competition is reduced, and the nature of this data use may attenuate reverse causality concerns.

Although a within estimator would alleviate most of the endogeneity concerns due to simultaneity, given the low within variation in the panel (that covers only one year), a fixed effects model at the app id level would not be an appropriate model as the fixed effects would completely capture the time-invariant factors. [17]

Various papers have employed the sellers' other app characteristics as instruments in their identification strategy (Comino, Manenti, and Mariuzzo, 2019; Cecere, Le Guel, and Lefrere, 2020). Indeed, developers' level characteristics are a crucial element of the app's success, and multi-app developers likely employ the same data strategy, update strategy, and pricing strategy to their apps.

I suggest utilizing a pooled OLS approach that includes developer-category dummy variables to address these factors. The first group of dummies considers the sellers' skills and other unobservable attributes, like their work methods (whether data-oriented or not) and managerial abilities. On the other hand, the second group of dummies considers time-invariant fixed effects that are specific to each category. This control is crucial because a data-savvy business model may not be viable in data-intensive sectors, like social networking, but can still be attainable in others, like Books. Therefore, it is essential to acknowledge the inherent differences between categories.

Therefore, the proposed econometric model is:[18]

[15] Using the number of worldwide downloads estimated by Sensor Tower may seem like the only option when cumulative installs are unavailable. However, this measure has its own set of problems since it has based on estimates and does not accurately reflect the number of users. Furthermore, the rating count can better proxy the number of actual users and estimate the number of satisfied users who may have the app installed. Additionally, professional data aggregators services like Sensor Tower, 42Matter, and AppAnnie, estimate the number of downloads starting from the rating count and the apps' rank weekly charts, further validating this proxy.

[16] We need to be aware that employing rating counts as a proxy of the installed base may favor free apps since more users download them hastily just to test them.

[17] As discussed in 1.4 the dependent variables exhibit very little within variation.

[18] The FE model has been chosen over RE after conducting the Hausman test, additionally cluster (submarket) specific fixed effects could not be introduced because the variable HHI varies only between c.

$$log(U2TU_{ict}) = \alpha_0 + \gamma_1 HHI_{ct} + \gamma_2 log(share_{ict}) + \gamma_3 log(ratings_{ict}) + \gamma_4 P^d_{ict} + \gamma_5 D^{inapp}_{ict} + \sum_{j=1}^{n} \delta_j Seller_j + \sum_{z=1}^{24} \theta_z Category_z + \mu_i + \varepsilon_{ict}, \quad (1.2)$$

therefore, the model implies that *Data Used to Track You* indicator is a function of concentration in the submarket (cluster), of the rating share (i.e., market share) of the app in that cluster, and the cumulative count of ratings, and a set of essential apps characteristics that capture apps' language, device, age rating group and maturity (distance from release date).

More specifically:

- $log(U2TU_{ict})$ is the dependent variable aggregated as explained in 1.2.2.
- HHI_{ct} is the traditional measure of concentration in cluster c at time t, and it was computed using the rating shares as a proxy for quantity and exploiting the community identification through network analysis of the 'suggested apps' data.[19]
  - In accordance to Kesler, Kummer, and Schulte (2019) results, concentrated submarkets exhibit higher data used. Therefore a positive sign shall be expected.
- $log(share_{ict})$ is the logarithm of the rating share as computed in 1.2.3.[20]
  - As this is a measure of the firm's market share, more powerful firms should exploit data more intensively (Dimakopoulos and Sudaric, 2018; Kesler, Kummer, and Schulte, 2019; Preibusch, Kübler, and Beresford, 2013). Therefore, the sign shall be positive.
- $log(ratings_{ict})$ is the logarithm of the cumulative count of ratings, used to split the effect of competition ($log(share_{ijt})$) from the firm size.
  - The expected sign is positive, as this is a measure of the market size of the firms. Larger firms may use more data.

[19] Sensitivity to an alternative and innovative methodology to define markets exploiting Natural Language Processing is in 1.6.

[20] Alternatively, I estimated the model with four market share dummies to look for non-linear relationships. Hence, five dummies were codified ($s_i \leq 5\%$, $5 < s_i \leq 20\%$, $20 < s_i \leq 40\%$, $40 < s_i \leq 80\%$,$80 < s_i \leq 1\%$) to express different classes of market shares and to isolate the last section that may be the one most likely biased by the imperfection of the 'You Might Also Like' measure of competition.

We formalize the hypothesis tested relative to market power, market concentration and data use as follows:

Hypothesis 1. *If data markups are present, we expect that concentration positively correlates with data to track individuals, so the estimated coefficient for HHI should be* $\gamma_1 > 0$.

Hypothesis 2. *Apps with higher market share and higher demand would have more market power and higher data markups, this would imply* $\gamma_2, \gamma_3 > 0$.

- P^d_{ict} represents a dummy variable equal to one if the app has an upfront price.
  - Given the results in Kummer and Schulte (2019), the expected sign is negative because of the substitution of revenue streams.[21]
- D^{inapp}_{ict} is an indicator variable of a business model based on in-app purchases, and it is a dummy equal to one if the app has at least one in-app purchase.
  - The expected sign may be negative because of the substitution of revenue streams, or it could be positive if data is an input for price discrimination and the two would be complementary.

We formalize the hypothesis tested relative to monetization and data use as follows:

Hypothesis 3. *The literature showed that in the Andoid ecosystem data and prices are substitute, if it is so also with Apple apps we expect* $\gamma_4 < 0$.

Hypothesis 4. *In-app purchases and data can be substitute if data is sold to advertisers and in this case we expect* $\gamma_5 < 0$*, or they can be complements if data is used for better price discriminate through various packages, and in this case we expect* $\gamma_5 > 0$.

- μ_i is composed by a set of app-specific and (often) time-invariant controls. The set of dummies and categories taken into consideration as controls are: *mac*, *iPad*, *iPhone*, *age_rating*, *#languages*, and *appmaturity*.
- *Seller* developers fixed effects, this captures the ability of the developer and developers' habits in extracting data. In the standard Pooled OLS, we control for the number of apps a seller has published. Additionally, I introduce the logarithm count of the number of apps by the developer.

[21] It would be interesting to check whether, in more concentrated markets, the firms do not differentiate and use both revenues from data and from prices (Casadesus-Masanell and Hervas-Drane, 2015)

  - This coefficient shall be positive in the Pooled OLS without seller FE and should lose significance when the seller dummies are introduced. Given that it varies over time when new apps are released, or old are dismissed, it does not drop out with seller FE.

- *Category* category fixed effects, to capture sector-specific time-invariant characteristics.

For completeness, I also report the coefficients of the POLS with only category dummies and the app id fixed effect model. All the proposed models have the log-log specification for market shares that have been chosen for two main reasons: the Ramsey Regression Equation Specification Error Test (*RESET*) results favored the log-log specification over the log-lev and lev-lev. Additionally, a common suggestion to reduce the problem of heteroskedasticity is to apply the logarithmic transformation (Fox, 2015). Furthermore, to correct for heteroskedasticity, we estimate the model with robust standard error, and we allow clustering of errors at the seller level.[22]

As for the choice between fixed effects and random effects models, the Hausman test firmly rejected the hypothesis of consistency of the random effect estimator. Therefore, the fixed effect model in (1.2) was estimated (Greene, 2003).

Before looking at the regressions' results, we discuss the specificities of the sample by providing summary statistics and preliminary inspection of the variables.

1.4 Sample Statistics

1.4.1 Summary statistics: data uses

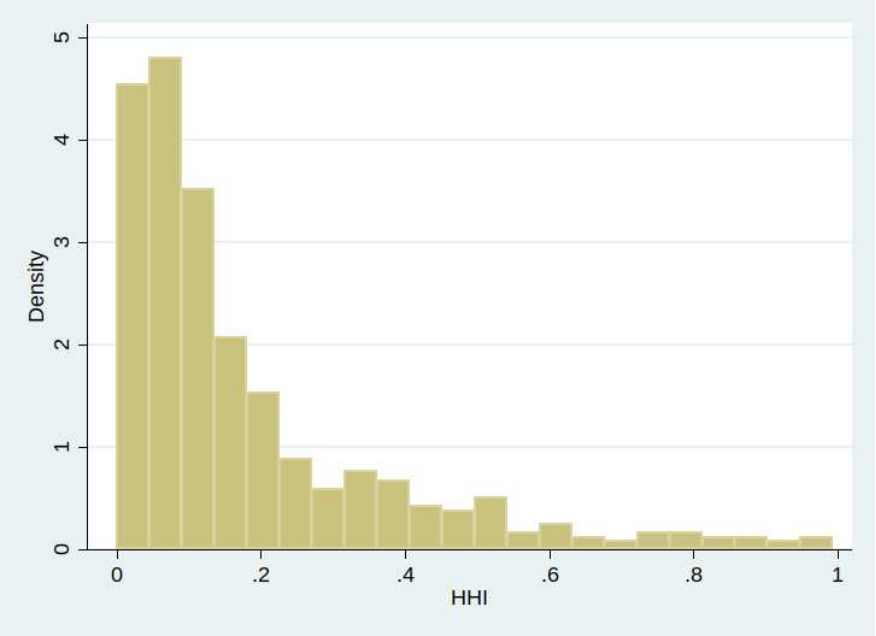
Table 1.1: Summary Statistics Data Uses

	count	mean	sd	min	max	skewness	kurtosis
Count of all the L2U data items by purposes	2410916	4.120	8.464	0	192	3.757536	27.82185
Count of all the U2TU data items	2410916	0.382	1.378	0	31	5.292713	41.00837
Dummy = 1 if any item U2TU	2410916	0.126	0.332	0	1	2.251939	6.071229
Dummy = 1 if any item L2U	2410916	0.378	0.485	0	1	.5030516	1.253061
Observations	2410916						

Table 1.1 shows the descriptive statistics and the distribution of the relevant indicators. The '*Data Linked to You*' variable is the sum of all the data items by each possible data use linked to the consumer profile. Therefore, the maximum of 192 is obtained by multiplying 32 possible data items collected over six possible data purposes. On the other hand, the link status *Data Used to Track* also represents a data

[22] Alternative clustering of standard errors that have been considered are: market level and category level and do not impact the significance of the main terms in the regressions.

Figure 1.3: HHI histogram



purpose. It does not include the other categories (advertising, app functionality, product personalization), and it can assume a value from a minimum of zero to 32. Given the considerable skewness of these two indicators, taking the logs of these variables is the standard practice to reduce skewness I also employ. [23] About 12% of the sample had at least one item tracked, while 36% and 40% of the sample had at least one item linked or not linked to the user profile, respectively. According to the linked status data, app functionality, and analytics were the most common purposes for data use.

1.4.2 Summary statistics: market definition and market shares

The market shares of apps, as proxied by the rating share and the Herfindahl–Hirschman index (HHI), have been computed in these clusters. Concerning the primary market definition methodology, the histogram in Figure 1.3 shows that the majority of markets are deemed competitive market (in 276 clusters $HHI \leq 1500$), a fraction has moderate concentration (in 131 clusters $1500 < HHI \leq 2500$) and more than a third is highly concentrated (in 205 cases $HHI > 2500$). The mean HHI along the clusters was .18 with moderate skewness (1.866) but relatively high kurtosis (6.3).

As a result, of the skewness of the rating count, the market share is highly skewed, and 99 % of the sample has an estimated share that is smaller than 5%. However, this concentration level is in line with statistics on the number of downloads, where

[23] A future extension may employ quintile regression that is particularly fit to work with skewed data. Given the skewness, as an additional test of the study's reliability, I analyzed the use of data items as dependent variables. I used an indicator variable to identify cases where at least one data item was tracked. This logistic regression model mirrored the pooled OLS model in Table 1.4 and confirmed the results from Section 1.5.

Table 1.2: Cluster Level Summary Statistics

YMAL Analysis							
	count	mean	sd	min	max	skewness	kurtosis
(mean) Market share	522	0.008	0.027	0	.5	13.01427	225.8422
(mean) hhi	522	0.183	0.196	0	.9930173	1.866231	6.321524
Mean number of firms	522	4618.6	8132.006	1	65908	3.33712	17.35601
Observations	522						
Description Analysis							
	count	mean	sd	min	max	skewness	kurtosis
(mean) Market share	3209	0.410	0.480	0	1	.3639937	1.179732
(mean) hhi	3209	0.424	0.482	0	1	.3073414	1.13825
Number of firms by cluster	3209	751.298	7758.296	1	185218	13.16196	208.6528
Observations	3209						

Note: these are the summary statistics after having collapsed the dataset at the cluster level and having retained the mean of the three variables. Therefore, every mean value refers to the mean of the cluster means.

SensorTower reported that in 2019 the top 1% of app publishers were responsible for 80 % of downloads. [24] Similarly, Bian, Ma, and Tang (2021) reports that the top 10000 apps account for 90% of downloads.

Other summary statistics for the panel dataset are reported in Table 1.2. There is a considerable difference in the number of clusters reported by the two procedures. Consequently, the mean HHI produced by the two market definition and their standard deviation is different, with more strictly defined markets having a higher concentration. Additionally, comparing the two market definitions reveals that the rating share is much less positively skewed in the case of the description analysis algorithm and with a much lower kurtosis.[25]

When using network analysis to define markets, there is a disadvantage because there is no clear metric to assess the clustering performance other than modularity. As a result, the reliability of the results across different clustering methods becomes crucial. [26]

Descriptive evidence of the positive association of data *U2TU* with market shares emerges from Figure 1.4, where there is an increasing trend between the market share category and the items reported in *U2TU*. However, the last class does not respect this trend, which may be due to a non-linear market power effect or a misclassification error

[24] SensorTower.com is a commercial data collector for aggregated statistics in the app sector.

[25] This is likely due to a more granular definition of markets that identified more clusters on the network from description analysis.

[26] It is worth noting that a test to evaluate the accuracy of the results that has been conducted and it is not reported in the article involves creating random clusters to demonstrate that the relationship between the clusters becomes insignificant as expected.

Figure 1.4: Log of *U2TU* by category of market share

of the modularity maximization algorithm. The misclassification is a possibility when the nodes have a low degree (like Apple products) and may end up in their cluster even though they have competitors. Thus, a further check that I do to isolate the misclassified apps or isolate non-linear effects is to use the categorized market share variable of Figure 1.4 and run the regression with four dummies. By breaking market shares into five categories, the misclassified apps tend to be the ones in clusters whose sum of ratings is less than 1000 and a market share (and HHI) close to 1.

1.4.3 Controls' summary statistics

Table 1.3 reports the summary statistics for the controls used in this study and other variables used to check the consistency of the sample with previous studies.

Firstly, we observe missing values for some of the variables, such as months from the release date (*m_old*), count of the languages, and apps' age rating that reduce the number of observations in the regressions. The missing values are entirely random due to the refusal of some HTML requests in the scraping process.[27]

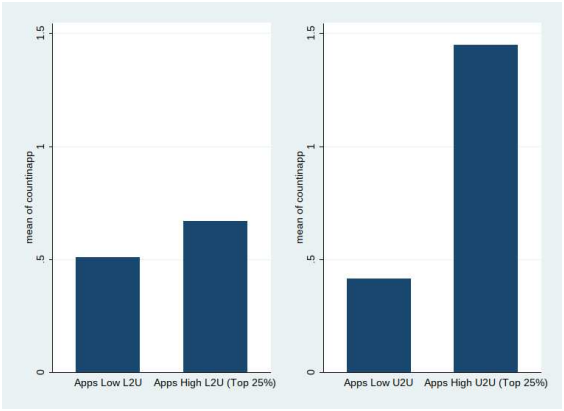
1.4.4 Summary statistics: Monetization

Concerning the monetization strategy only 7% of the sample has a positive price, and 14% of the sample has in-app purchasing options. There is consistent variability

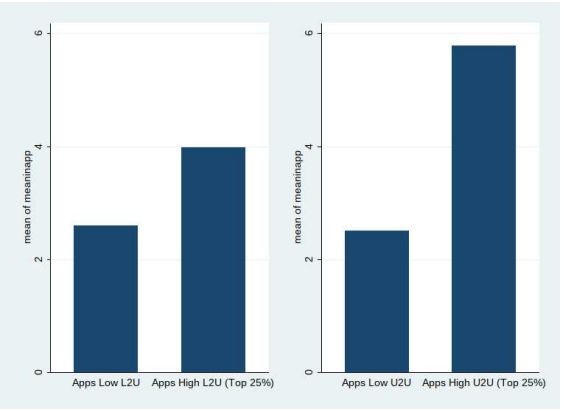
[27]To scrape the apps, a proxy service has been used with IP listed in the US. However, the proxy failed some requests that produced the missing values

in the purchasing options offered in the number of packages and the average price of the package, and one interesting future research avenue would be to explain the impact that data has on the ability of apps to price discriminate. I contribute to this theme by providing descriptive pieces of evidence: Figure 1.5a and 1.5b show that the last quartile of the sample in both the variable *L2U* and *U2TU* has a higher level of in-app purchasing options and a higher average price of in-app purchases with respect to the other 75% of the distribution. Furthermore, there is a consistent (and statistically significant) difference among the top 25% apps for data *L2U* from the top 25% apps for *U2TU* use. Moreover, further inspection of the few apps that changed the privacy

Figure 1.5: In-app purchases indicators for last quartile of L2U and U2TU vs. first three quartiles



(a) Mean number of in-app purchases packages



(b) Average price of in-app purchases packages

Note: The figure splits the sample with the top 25% of data use *L2U* (the left subplots in each figure) and of *U2TU* (the right subplots in each figure) against the remaining part of the respective distribution. Apps that use more data have an average number of packages available for purchase: the left panel shows the mean for the number of packages available for *L2U* (the left subplot) and *U2TU* (the right subplot). Apps with more data used have a higher average price (three times the sample averages for *U2TU*).

panel (*U2TU*) shows significant differences in terms of rating count and market share. On average, the apps that changed the summary have a rating count and market share that is twice those of the apps that did not change it. This fact would indicate a market power manifestation in surplus extraction for this category. A more in-depth descriptive analysis of the apps that changed the privacy indicator *U2TU* is provided for the interested reader in A.3.

1.4.5 Summary statistics: Updates

The data is also rich regarding updates information by providing the history of the last 25 updates of the apps with their associated date and topic. By dividing

the number of updates by the update range expressed in months, an indicator for the number of updates per month was obtained. The average app in the sample updates the app every 50 days, which is remarkably similar to the figure found in Comino, Manenti, and Mariuzzo (2019). The apps that have higher values of the privacy indicators also have a higher number of updates. However, the app age may be driving this correlation since older apps are more likely to have more updates and, simultaneously, have a more extensive installed base that would motivate higher profitability of data trades. Therefore, regression analysis to separate the effect of data on updates would be needed. A short extension on this theme is proposed in A.4.

Table 1.3: Summary Statistics

	count	mean	sd	min	max	skewness	kurtosis
Count of ratings (in thousands)	2410916	2.082	95.290	0.000	28500.000	156.890	33480.235
Average stars, NA if rating_count==0	1323377	4.217	1.053	1.000	5.000	-1.593	4.769
Price dummy, =1 if price>0	2410916	0.072	0.258	0.000	1.000	3.314	11.984
Dummy variable for in-app purchases	2410916	0.141	0.348	0.000	1.000	2.058	5.237
Numeric count of number of packages	2410916	0.546	1.822	0.000	20.000	4.160	20.994
Avg. Price of the inapp purchases	2410916	2.924	19.286	0.000	999.990	21.154	701.232
N. apps by seller and wave	2410916	45.742	321.573	1.000	3991.000	10.877	125.712
Months from release date	2401033	40.185	33.600	0.000	171.000	1.199	3.794
Age Rating (PEGI), 4+ 9+ 12+ or 17+	2410901	6.070	4.364	4.000	17.000	1.798	4.541
Dummy variable, =1 if mac version exist	2410916	0.775	0.418	0.000	1.000	-1.314	2.726
Dummy variable, =1 if ipad version exist	2410916	0.653	0.476	0.000	1.000	-0.644	1.415
Dummy variable, =1 if iphone version exist	2410916	0.981	0.135	0.000	1.000	-7.118	51.666
Numeric count of the languages of the app	2335136	3.588	6.889	1.000	141.000	4.565	32.237
Number of updates per month	2410916	0.630	0.895	0.000	25.000	3.896	34.955
Numeric count of the updates done (capped at 25)	2410916	9.759	8.431	1.000	25.000	0.781	2.154
Observations	2410916						

1.4.6 Other controls

Despite the table indicating an average age of the app of 40 months, this does not consider that the panel is unbalanced, and we observe entrants for only a couple of periods. As a result, the average maturity of the apps in February 2022 was 36 months. Over the whole panel, the developer's average number of apps is 45.72, with a substantial standard deviation. Considering that some developers have almost 4000 apps, controlling for developers' unobservable characteristics becomes paramount given this high variance. Almost all the sample has a version for the iPhone (98%), while around 70% also has a version for the macOS operating system. Given that multi-device apps may provide richer data for tracking the users, it is interesting to check whether multi-device apps collect more or less data.

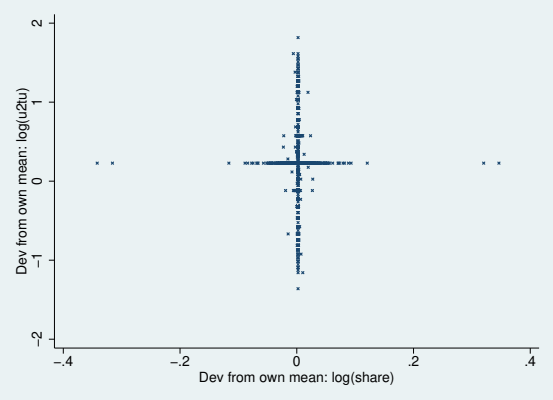
1.4.7 Inspection of within variation

The panel's summary statistics indicate that the within standard deviation for the $U2TU$ indicator is low compared to the between variation, with values of 0.09 and

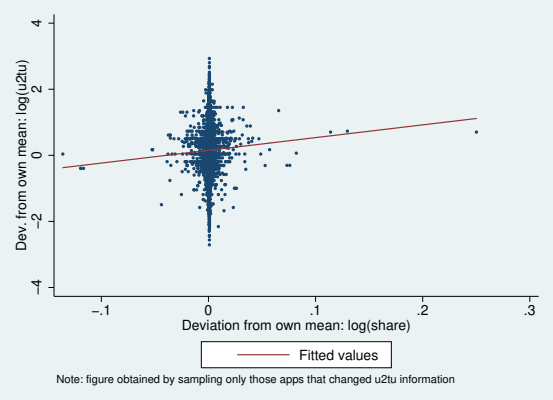
0.44, respectively. This trend is even more pronounced with market shares, which have values of 0.001 and 0.011, respectively.

In Figure 1.6a, I plot the deviation from the apps' own mean for the $U2TU$ indicator against the log deviation of market share for the whole sample. While data strategy is a long-term strategic variable, the panel used in this study covers only one year. Therefore, it is natural that the within-variation in these variables is not significant. Consequently, it is difficult to observe a significant trend between the deviations of the two variables.[28] Although the left panel of the figure confirms that given the short observation period, most of the variability is between apps, by focusing on the apps that showed a change in $U2TU_i$ (the right sub-figure 1.6b), a weak positive correlation emerges.

Figure 1.6: Inspection of within variation for main variables



(a) Deviation from the mean for $log(U2TU)$ and $log(share)$



(b) Deviation from the mean for $log(U2TU)$ and $log(share)$ for those that changed $U2TU$

Furthermore, the same weak positive correlation emerges also if we look at the deviation from the mean of the app maturity (months from release) and the deviation from the mean of $U2TU$ (see Figure A.3).

In the next section, the results from the main regressions are reported.

1.5 Results

1.5.1 Main model: the impact of market power on *Data Used to Track You*

Table 1.4 reports the results of the regression analysis: the first column shows the results of standard OLS with category dummies, the second reports the model expressed in (1.2), and the third displays the same model with categorized market shares to look for non-linear effects. The fourth and fifth columns mirror the model in

[28] I do not have access to Apple App Store's historical data, and I am currently collecting data quarterly to expand the dataset to cover multiple years and repeat the analysis over a longer period.

the second and third columns, with the difference that the fixed effects are at the app id level like in Kesler, Kummer, and Schulte (2019). Differences in the sample from the total number of observation arises from the automatic drop of singletons when estimating the regression. Using singleton would lead to artificially underestimating the standard errors and artificially increasing significance (Correia, 2015). Additionally, given the concern of heteroskedasticity, heteroskedastic-robust standard errors are provided in parenthesis.

Market concentration effects The regressions with Pooled OLS with category FE and the columns of the category-developer FE model show a small but significant positive association of concentration and *Data Used to Track You*, while, when introducing app fixed effects, the estimated coefficient unsurprisingly becomes insignificant due to the low within-variation of these variables as reported in 1.6a. Even when the magnitude is at the largest estimate, the magnitude of the HHI estimated coefficient seems to be neglectable: with a complete transition from perfect competition to monopoly, the parameter estimated in column (1) would imply an increase in *Data Used to Track You* of about 2%. Moreover, this effect is not robust to different specifications and market definitions, and the positive association fades when running robustness tests on the resolution parameter and alternative forms of market definition.

Market power effects However, the market share proxy coefficient is highly significant in both the first two models and loses significance when moving to app-level fixed effects. However, the loss in significance can be attributed to the low within-variation in the sample and the short time dimension of the panel. The app fixed effect model explains 95% of the variance, and correspondingly it is expected that (almost) time-invariant factors lose statistical significance and get lower estimated coefficients. Additionally, some endogeneity may come from the measurement error of market shares. In fact, by breaking the market share into categories, there either is a non-linear effect or the noise of the measure for market share may impact the coefficient of the last category. The market definition may be particularly problematic when the YMAL network is skewed with lower degree nodes (like the entrants/new apps). In this case, these apps form communities with low impact for their market size (demand) but very high market shares that tend to be unitary. This would explain the high and positive significance of the fourth category ($0.4 < s_i \leq 0.8$) in column (5) (model with app-level fixed effects) and the non-significance of the last one. Concerning the interpretation of these results, we infer from the log-log model with category and seller fixed effect that the relationship between market power and *Data Used to Track You* is inelastic: an increase of market share by 1% would translate to an increase of about 0.43% in data uses. Although the magnitude varies consistently across

Table 1.4: Selected coefficients for model on *Data Used to Track You*

	POLS	Developers Dummy		App FE	
Dep. Var: log(u2tu)	(1)	(2)	(3)	(4)	(5)
HHI	0.024***	0.005*	0.006*	-0.003	-0.003
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
Log(share)	0.703***	0.443***		-0.067	
	(0.047)	(0.041)		(0.072)	
Categorical market shares (baseline x≤ 0.5)					
$0.05 < x \leq 0.2$			0.045***		-0.000
			(0.008)		(0.008)
$0.2 < x \leq 0.4$			0.062***		0.028
			(0.017)		(0.019)
$0.4 < x \leq 0.8$			0.330***		0.081**
			(0.033)		(0.030)
$0.8 < x \leq 1$			-0.005		-0.109
			(0.042)		(0.156)
Log of rating count	0.033***	0.019***	0.019***	0.015***	0.015***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Price dummy, =1 if price> 0	-0.140***	-0.210***	-0.210***	-0.027***	-0.027***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Dummy variable for in-app purchases	0.182***	0.047***	0.047***	0.043***	0.043***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Log(N. apps) by seller and wave	0.020***	-0.020***	-0.020***	-0.013***	-0.013***
	(0.000)	(0.002)	(0.002)	(0.001)	(0.001)
App age Categories (baseline 0-12m/o)					
Young (13-21m/o)	0.003***	-0.000	-0.000	0.005***	0.005***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Mature (22-37m/o)	-0.016***	-0.001	-0.001	0.016***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Very Mature (38-66m/o)	-0.023***	0.000	0.000	0.029***	0.029***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Veteran (67-121m/o)	0.004***	0.009***	0.009***	0.047***	0.047***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.044***	0.166***	0.165***	0.141***	0.141***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Fixed Effects	|	|	|	|	|
Category	✓	✓	✓	✓	✓
Developer	|	✓	✓	|	|
App	|	|	|	✓	✓
Observations	2317525	2317525	2317525	2317525	2317525
R^2	0.121	0.847	0.847	0.955	0.955

Note: this regression controls for device supported (mac), age rating of the app (+4,+9,+12,+17) category fixed effects, count of languages. The set of dummies used in each model is in the last section of the table. All models have been estimated through the Stata command regdfe that automatically drops singletons to ensure the standard error is not underestimated. The sample has been restricted from the original sample to ensure that each estimated model had the same number of observations. Variations in the sample did not substantially modify the coefficients' magnitude, significance, and sign. Significance levels are: *** p<0.001, ** p<0.01, * p<0.05

specifications, this effect's sign and significance are robust to market definition and resolution parameters, albeit decreasing in this last element.

We state this first set of results in the following proposition:

Proposition 1. *The regression analysis results do not fully support Hypothesis 1, as the significance of the estimated coefficient γ_1 is not robust to changes in the specification and market definition. Furthermore, even when the estimated coefficient is significant, its impact on the dependent variable is negligible. On the other hand, the results support Hypothesis 2: as the estimated γ_2 is positive and significant across multiple specifications, the correlation of market power's with the self-reported 'Data Used to Track You' indicator may be the result of a market power effect.*

In the following paragraph, we report some ancillary unexpected results that may offer further support to Proposition 1.

Apps' Maturity Introducing a categorical variable encoding the quintiles of the months since the app's release provides intriguing insights into the relationship between an app's maturity and *Data Used to Track You*. This effect, although not particularly large in magnitude, is robust to the inclusion of sellers' dummies. This coefficient reveals a novel aspect of *Data Used to Track You*: the oldest apps tend to track consumers more across apps and possibly sell more consumer data to third parties, even after controlling for developer-fixed effects and the number of apps available by the same seller. Interestingly, the category of the app's age is significant across all specifications and robust to the app-level fixed effects. Therefore, older apps associate with higher *Data Used to Track You* with respect to very young apps. The fact that the effect reinforces when introducing apps' level fixed effects indicates that the aging of the apps in this short panel was already relevant to identify the effect of a data markup. Finally, while the time-invariant component of the app is captured by the within estimation, the app maturity indicator (that varies over time) that becomes more prominent and statistically significant may indicate that there is a dynamic increase of the data markup associated with the aging process of the firms of the sample.

Multiple explanations are possible, and more research is needed on this correlation. Theoretically, as the firm ages and grows, the stock of data becomes larger, richer, and more informative. Consequently, the marginal revenue from selling data may be increasing, and the older the app and the more the developer tends to substitute the price source of revenue with other data-driven sources (ads, sales of information to data brokers). A popular alternative explanation is one of data barriers: apps that survived are those that employed more data to track consumers in the first place by raising barriers and using data strategically not only to ensure the survival of their

business but to push out rivals that were not using data.[29] A final potential cause of this result that could not be excluded is the measurement error in the dependent variable, further discussed in the 1.6.3.

Monetization One intriguing finding is that the coefficient for the price dummy is negative, while the coefficient for in-app purchases is positive. This suggests that while upfront app prices and data are substitutes, in-app purchases are linked to "*Data Used to Track You.*" Additionally, these effects hold across various specifications with similar estimated magnitudes. This result supports the theory that data used in this manner is utilized for price discrimination. As stated in the descriptive section and confirmed in the regression, "*Data Used to Track You*" may be connected to "freemium" business models that can engage in price discrimination by lowering quality through advertising and then offering ad-removal packages. While this analysis is not covered in this paper, it presents an exciting avenue for future research, especially given the wealth of data on this topic. Let us formalize the results in the following proposition:

Proposition 2. *The resulting correlation support Hypothesis 3. Data and prices are substitutes in the Apple ecosystem, as the estimated coefficient γ_3 is negative and highly significant across multiple specifications and market definitions. Concerning Hypothesis 3, the positive sign of γ_4 suggests that in-app purchases and data are complements, and apps may use data to maximize their profits through price discrimination.*

Other controls It is interesting to notice that differently from Kesler, Kummer, and Schulte (2019), the log of rating count, which should capture the size or demand of the firm, remains positive and significant and reinforces the leading market share effect. However, the full effect is still below the unitary elasticity.

The log number of apps of the developer has only been introduced as a control because otherwise, when developer fixed effects are not included, there would be omitted variable bias: it is expected that developers that have more apps also use more data to track consumers across their own (different) apps. It is also expected that the significance of this term disappears in the fixed effects models. However, the fact that

[29] It is crucial therefore to study entrants' behavior more in detail in future research. An example of the data-driven exclusionary practice was the attempt of Google to exclude rivals from accessing users' Big Data in European Commission, Case AT.40099 – Google Android: *"In addition to allowing Google to maintain and deepen its dominance in online advertising, its data collection has allowed Google to entrench its dominance in search. As the EC is well aware, the advantage conferred to Google by its scale in data – combined with the anti-EN 97 EN competitive conduct Google employs to protect its position – has raised insurmountable barriers to entry in the markets for general search and in particular specialized search services. [...] In addition to giving Google an advantage in search and online advertising, the data Google collects gives it an advantage in optimizing its mobile (and PC) services such as YouTube and Maps, as well as in predictive technologies such as Google Now. For example, one way Google can gain competitive insight into user behaviour is to understand which apps are installed, or removed, by users on its platform."* Oracle Statement in the Google case available at https://ec.europa.eu/competition/antitrust/cases/dec_docs/40099/40099_9993_3.pdf

it becomes negative when introducing fixed effects was not expected and could capture the launch of new apps by the same developer. This is because the developer fixed effect would capture the time-invariant characteristics, but the log(number of apps) may vary over time precisely when the developer dismantles old apps or launches new apps.

The other regressors introduced in the model to control for the app (mostly) fixed characteristics in the model (2) are the age rating of the apps, which show the comforting fact that the apps that also target kids (+4) are those that track users less, a dummy for the presence of the macOS version and a count of the languages of the app that shall capture a demand effect and confirms the positive demand effect on *Data Used to Track You* proxied by the rating count. These results are available in the full Table A.1 in the Appendix.[30]

1.6 Robustness

1.6.1 Sensitivity to cluster resolution

Table A.1 shows the relevance of the resolution parameter for the result. Given that the Louvain Algorithm is a hierarchical clustering and the resolution parameter gives the level at which the algorithm stops, the sensitivity of the results to this parameter shall be discussed. In the table in the Appendix, I show that the result is not dependent on the level of resolution chosen as long it is not too large. To be conservative on the effect of market power, the effects from the model reported in 1.4 that have been selected (by validating the clusters by hand) are the smallest of the series. The table shows a clear negative relationship between the resolution parameter and the effect of market share on data used to track. However, a resolution parameter larger than the one reported tends to create a high share of clusters with apps with low aggregate demand but extremely high market share (mostly one). This artificially decreases the significance of the market share coefficient.

1.6.2 Other market definitions

Similarly to Kesler, Kummer, and Schulte (2019), I tested alternative proxies of competition. Firstly, the market share results do not carry over when considering only the "radius around the plant" measure built by taking the market share of one app among the vector of similar apps. Secondly, I propose a sensitivity test based on a market definition measure that does not use the You Might Also Like (YMAL)

[30] A non-reported test was done with a logit model on the dummy equal one if any item is reported in the $U2TU$ privacy label. It fully confirmed the results of the linear regressions, and it is available upon request.

network. Given that the YMAL section may fail to provide similar apps for new and smaller apps and the fact that Apple does not provide the YMAL section for its apps, such as iMessage or Apple Music, the distribution of communities obtained through modularity maximization may not capture the categorization of new apps correctly. Therefore, I enrich this with an analysis of the description text similarity through a Natural Language Processing analysis that exploits the textual descriptions of the apps and network analysis to identify apps' submarkets (similarly to Hoberg, Phillips, and Prabhala (2014) and Pellegrino (2023) and Hoberg and Phillips (2010)).

The results and detailed methodology are reported in Appendix A.2.2, and Table A.2 shows that although the magnitude is lower, the significance and sign of the effect of market share proxy is robust to this new market definition.

1.6.3 Endogeneity concerns

Endogeneity concerns are addressed in Appendix A.1, and here I summarize the main issues treated in the Appendix. A first concern arises from the potential reverse causality of data and market share: the focus on *Data Used to Track You* reduces the concerns for reverse causality because this data use has the largest surplus extraction term. Therefore, the firm that extracts more data is less likely to increase the utility offers to consumers. Consequently, there is a higher likelihood that data do not impact the reaction function in the firm's utility (De Cornière and G. Taylor, 2023).

A second concern may arise from data being an input in the updating process. On this issue, a preliminary regression analysis in Appendix A.4 shows that this indicator has a negative conditional correlation with the count of updates and the probability of a version change, and this would suggest that $U2TU$ data is not directly used as an input in the updating process.

Nonetheless, a third potential bias may derive from simultaneity in choosing the data items in each data use panel. Applying app-level fixed effects and the robustness of the categorical market share variable in this regression alleviates the concern for simultaneity bias driven by economies of scale in data collection. In addition, future dataset expansion and repeating the analysis over a longer time period may further reduce concerns for simultaneity.

Sensitivity to different specifications

The sign and significance of the main effect of market share are robust to a change in specification from log-log to log-lev and lev-lev. Moreover, this remains significant

even considering as dependent variable the following:

$$\text{Data Mark-up} = \begin{cases} u2tu - l2u \;\; \text{If} \;\; l2u \leq u2tu \\ 0 \;\; \text{Otherwise} \end{cases}$$

This definition exploits the data items that are unique to the $U2TU$ section. This specification of the dependent variable has the advantage of being more robust to simultaneity bias and the results with this regression were excluded as they were fully in line with the main results presented in Table 1.4.

Measurement error in dependent variable

Both dependent and explanatory variables used in this study are proxies for the variables of interest and may be subject to measurement error. While the measurement error for the explanatory variables is discussed in the market definition robustness tests, we discuss the possible impact of measurement error for the dependent variable by basing the discussion on Wooldridge (2015).

Regarding the dependent variable, it was impossible to address the concern of strategic reporting of privacy labels in the present research.[31] Generally, the estimates obtained through POLS will still be unbiased if the reporting error is statistically independent of each explanatory variable (for example, if every developer cheats slightly or is randomly distributed). However, this will not be true if the reporting error correlates with the estimated regressors. Indeed, if there is no sanction for misrepresenting privacy policies, developers do not have many incentives to report correct information, and the possibility that everyone cheats is sensible. Therefore, absent Apple's intervention, the reporting error is i.i.d. within the sample, and the estimates are unbiased.[32]

However, Apple's declared that developers must update their privacy labels when found "guilty", and this intervention may influence the statistical independence of the reporting error in multiple ways. The likelihood that Apple finds a developer guilty can be modeled as a function of the duration of the deceptive behavior of the app, the number of consumers/popularity of the app, and the lobbying activity of the developer. Understanding the correlation of market power with these three elements is crucial to

[31] See for example privacy labels article.

[32] Consider the simple regression model that satisfies the Gauss-Markov assumptions:

$$y^* = \beta_0 + \beta_1 x_1 + u$$

where $e_0 = y - y^*$ is the measurement error. If we only observe y and we write a model for it:

$$y = \beta_0 + \beta_1 x_1 + u + e_0,$$

we see that we need to have e_0 independent of x_1 to have a unbiased estimate for β_1.

determine the direction of the bias.

Firstly, Apple's initiative is relatively recent, and we can assume that the duration of the deceptive behavior is similar among high and low market share firms. Secondly, powerful firms are more likely to get caught by Apple because of popularity. Then, while they would be constrained to their actual value, smaller apps may still get away with deception and keep $U2TU_{TRUE} > U2TU_{REPORTED}$. If this is the case, the impact of market power on U2TU (represented by γ_2 in the model's equation) may be exaggerated. Thirdly, larger firms (and multi-app developers) also have more resources for lobbying and could be able to invest to capture 'regulators'. In this case, the bias would go in the opposite direction, and the estimated relationship would be understated. In conclusion, we need more information about the correlation between market power and the under-reporting phenomenon to provide a clear direction for the bias. Finally, at this stage, Apple's intervention has been minimal. Therefore, this article estimates the effect of competition on the self-reported app's privacy policies under the assumption that the practice of misreporting is still statistically independent from the regressors in the equation.

1.7 Conclusions

This study investigates the correlation between data and market power, utilizing a unique dataset obtained from the Apple App Store. The impact of market power on data usage was estimated by meticulously collecting and analyzing data on all available apps from the online iTunes catalog. Market definitions were determined using a network science technique based on the modularity of the network of similar apps provided by Apple, and an alternative market definition was tested using a network of description cosine similarities to ensure the robustness of the results. Despite data limitations, rating count was utilized as a proxy for downloads and to build market shares and concentration indexes. Differently from the previous literature, the intensity of data usage to track individuals was the focal point of the analysis rather than the number of permissions required by an app. This indicator was built from the privacy labels provided by app developers, and it alleviates concerns for reverse causality due to the high privacy concerns elicited and the high surplus extraction role of data in this link status. A Pooled OLS model was utilized for the primary analysis, introducing developer and category dummies accounting for time-invariant factors.

The estimated marginal impact of HHI on *Data Used to Track You* indicator is negligible, and the effect is not robust across specifications. On the other hand, apps' market share proxies have a significant and positive effect on data $U2TU$. The magnitude suggests an inelastic relationship between market share and data intensity, with the log-log form suggesting that an increase in market share by 1% is associated with

about 0.4% more data items. Instead, the categorized shares show that the effect may be non-linear, and apps that are dominant and quasi-dominant firms that fall within the 40-80% use about 33% more items to track consumers. While the magnitude drops consistently (to 8%) in the case of the app's fixed effect, the significance of this result carries over. This result has been obtained by controlling for the app's maturity (distance in months from release date), and it emerged that this control variable becomes highly significant and more impactful in the app fixed effects case. This suggests that the amount of data *U2TU* increases as apps age. It remains to understand whether this represents a shift in business model or whether there are selection effects at stake that may be motivated by the anti-competitive use of data.

This article contributes to the existing literature by examining the extent to which various apps use data, allowing for the identification of their data strategy. This approach is advantageous as it distinguishes between data collection purposes that may increase market share and those that may have a small or negative impact, thereby decreasing the risk of reverse causality in estimates.

Furthermore, this focus on data usage, rather than the amount of data collected, adds to the economics of app literature by highlighting the correlation between "*Data Used to Track You*" (*U2TU*) and in-app purchases, app maturity, and updates.[33] The apps with higher values of *U2TU* also tend to use more in-app purchasing options with a higher average price and are marginally older than those with more data linked to user profiles (*L2U*). This trend is consistent across different perspectives, with those apps that are more active in changing their *U2TU* panel and introducing new items having three times the number of packages compared to the sample average.

Additionally, our regression analysis confirms that using data to track consumers complements the presence of in-app purchases, while the prices and data are interchangeable. This finding is novel with respect to the literature and fully explaining its causes was beyond the scope of this article. Further investigation is required to explore why the correlations between data and upfront prices and data and in-app purchases have opposite signs. The descriptive evidence indicates that apps may offer various qualities in the market to enable price discrimination through in-app purchases. These apps can price discriminate by collecting data from different sources and sharing them with third parties. A structural model in this field would help determine the welfare effects of data when this type of price discrimination occurs.

By merging two methodologies, namely network modularity maximization (as in Kesler, Kummer, and Schulte (2019)) with Natural Language Processing (NLP) and text analysis (as in Hoberg and Phillips (2010), Hoberg, Phillips, and Prabhala (2014), and Leyden (2018)), this article provides a methodological contribution and an alternative way to define digital markets that are particularly promising in applied work

[33]The analysis of updates is in the Appendix as it is still in an early stage.

when text data is available.

To sum up, this essay shows that market shares positively correlate with the data used to track consumer behavior, but this correlation is smaller than anticipated. The findings are consistent across different specifications, but the limited variation within the panel makes it challenging to maintain statistical significance for all market share categories when using a within estimator. Therefore, this is the first limitation of the study, which could be addressed by analyzing a more extended panel in future research.

Moreover, the identification strategy in this article is based on theoretical assumptions regarding the impact of data on competition, which may be restrictive. While these assumptions are plausible, it was impossible to test for simultaneity in the choice, which could bias the market share estimates. Only the coefficients of the categorical market share, which remain significant with app fixed effects, and the effect of app age on data use intensity are robust to simultaneity bias. These were obtained through panel fixed effects that captured within-variation and eliminated time-invariant factors.

Therefore, at this stage the results must be read with the appropriate caution and given the many sources of endogeneity the estimated coefficients represent conditional correlation that do not imply causality. Future investigation exploiting exogenous changes, such as the introduction of other app stores, app recategorizations, or governmental-imposed bans of apps may provide an instrument to confirm these estimates.

Another limitation of the study is the measurement error and associated attenuation bias of the proxies employed for market concentration and the self-reported nature of the privacy labels.

On the first concern, although the correlation between rating count and downloads is solid, free apps may be overrepresented in some categories. Therefore, this measure may overestimate the market share of free apps that receive more downloads, reviews, and uninstalls. Focusing on an app's installed base is challenging, but it may be a better proxy for market power in future research. One way to estimate this measure could be to count only ratings with at least three stars.

Finally, the research could not address concerns about strategic reporting of privacy labels. If the reporting error is independent of explanatory variables, estimates from POLS will be unbiased. However, the estimates will be biased if the error is correlated with estimated regressors. Although the hypothesis of statistical independence is reasonable without platform's intervention, Apple's requirement that developers update privacy labels when found guilty may affect the statistical independence of reporting errors. We need more information on market power correlation to determine the bias direction. Nevertheless, with the label's introduction being so new, the article assumes that the impact of Apple's intervention is minimal and misreporting is still

independent of regressors.

Future research This study did not explore the impact of innovation on data, which is an essential aspect. Innovation is usually seen in new apps or updated versions of existing ones. With approximately 2000 new apps being submitted daily, it is impossible to keep track of all of them since the catalog is not updated that frequently. As a result, I only observed a fraction of the entrants, about 65000 in a year. These could either be the ones that successfully passed the initial developmental stages and complied with all the required rules and regulations, or they could be entirely random. To better understand the effects of data on competition and how it alters developer incentives, future research should investigate the connection between a developer's likelihood of launching new apps and the data they collect, also through different apps.

I only use updates in the Appendix to test for potential feedback loops. Incremental updates are essential for non-game app developers, so the analysis of updates in the Appendix can lead to new research on the positive effects of data. One potential area of research could be analyzing app update behavior through survival analysis, which could answer questions about the impact of consumer data on update quality and the length of time an app can survive without updates. This could also help determine if reducing the number of updates while increasing their quality could increase the "buzz effect" reported in previous studies. Additionally, categorizing updates as bug fixes, feature expansions, or pricing updates (in a way similar to Leyden (2018)) could help expand our understanding of the positive effects of data on innovation.

References

Acquisti, A., L. Brandimarte, and G. Loewenstein (2015). "Privacy and human behavior in the age of information". In: *Science* 347.6221, pp. 509–514. ISSN: 10959203. DOI: 10.1126/science.aaa1465.

Acquisti, A., C. Taylor, and L. Wagman (2016). "The economics of privacy". In: *Journal of economic Literature* 54.2, pp. 442–92.

Acquisti, A. and H. R. Varian (2005). "Conditioning prices on purchase history". In: *Marketing Science* 24.3, pp. 367–381.

Bergemann, D. and A. Bonatti (2019). "Markets for information: An introduction". In: *Annual Review of Economics* 11, pp. 85–107.

Bergemann, D., A. Bonatti, and T. Gan (2022). "The economics of social data". In: *The RAND Journal of Economics* 53.2, pp. 263–296.

Bian, B., X. Ma, and H. Tang (2021). "The supply and demand for data privacy: Evidence from mobile apps". In: *Working paper available at SSRN*.

Blondel, V. D. et al. (Oct. 2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008.

Calzolari, G. and A. Pavan (2006). "On the optimality of privacy in sequential contracting". In: *Journal of Economic theory* 130.1, pp. 168–204.

Casadesus-Masanell, R. and A. Hervas-Drane (2015). "Competing with privacy". In: *Management Science* 61.1, pp. 229–246.

Cecere, G., F. Le Guel, and V. Lefrere (2020). "Economics of free mobile applications: Personal data and third parties". In: *Available at SSRN 3136661*.

Comino, S., F. M. Manenti, and F. Mariuzzo (2019). "Updates management in mobile applications: iTunes versus Google Play". In: *Journal of Economics & Management Strategy* 28.3, pp. 392–419.

Correia, S. (2015). "Singletons, cluster-robust standard errors and fixed effects: A bad mix". In: *Technical Note, Duke University* 7.

De Cornière, A. and R. De Nijs (2016). "Online advertising and privacy". In: *The RAND Journal of Economics* 47.1, pp. 48–72.

De Cornière, A. and G. Taylor (2023). *Data and competition: A simple framework*. Tech. rep.

Dimakopoulos, P. D. and S. Sudaric (2018). "Privacy and platform competition". In: *International Journal of Industrial Organization* 61, pp. 686–713.

Etro, F. (2021). "Device-funded vs ad-funded platforms". In: *International Journal of Industrial Organization* 75, p. 102711.

European Commission (2019). *Special Eurobarometer 487a: The General Data Protection Regulation*. Tech. rep. March, p. 104. DOI: 10.2838/579882. URL: http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/86886.

Farboodi, M. et al. (2019). "Big data and firm dynamics". In: *AEA papers and proceedings*. Vol. 109, pp. 38–42.

Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications.

Fudenberg, D. and J. Tirole (2000). "Customer poaching and brand switching". In: *RAND Journal of Economics*, pp. 634–657.

Ghose, A. and S. P. Han (2014). "Estimating demand for mobile applications in the new economy". In: *Management Science* 60.6, pp. 1470–1488.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education.

Hoberg, G. and G. Phillips (2010). "Product market synergies and competition in mergers and acquisitions: A text-based analysis". In: *The Review of Financial Studies* 23.10, pp. 3773–3811.

Hoberg, G., G. Phillips, and N. Prabhala (2014). "Product market threats, payouts, and financial flexibility". In: *The Journal of Finance* 69.1, pp. 293–324.

Kesler, R., M. Kummer, and P. Schulte (2019). "Competition and Privacy in Online Markets: Evidence from the Mobile App Industry". In: *ZEW Discussion Paper* No. 19-064.

Kummer, M. and P. Schulte (2019). "When private information settles the bill: Money and privacy in Google's market for smartphone applications". In: *Management Science* 65.8, pp. 3470–3494.

Leyden, B. T. (2018). *There's an app (update) for that*. Tech. rep. mimeo.

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

Newman, M. E. (2006). "Modularity and community structure in networks". In: *Proceedings of the national academy of sciences* 103.23, pp. 8577–8582.

Norwegian Consumer Council (2018). *Deceived By Design*. Tech. rep. URL: `https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf`.

Pellegrino, B. (2023). "Product differentiation and oligopoly: a network approach". In:

Preibusch, S., D. Kübler, and A. R. Beresford (2013). "Price versus privacy: An experiment into the competitive advantage of collecting less personal information". In: 13.4, pp. 423–455.

Prüfer, J. and C. Schottmüller (2021). "Competing with big data". In: *The Journal of Industrial Economics* 69.4, pp. 967–1008.

Schmalensee, R. et al. (1989). *Handbook of industrial organization*. Vol. 3. Elsevier.

Taylor, C. R. (2004). "Consumer privacy and the market for customer information". In: *RAND Journal of Economics*, pp. 631–650.

Tsai, J. Y. et al. (2011). "The effect of online privacy information on purchasing behavior: An experimental study". In: *Information systems research* 22.2, pp. 254–268.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Yin, P.-L., J. P. Davis, and Y. Muzyrya (2014). "Entrepreneurial innovation: Killer apps in the iPhone ecosystem". In: *American Economic Review* 104.5, pp. 255–259.

Chapter 2

Data Externalities and Vertical Differentiation in Digital Markets: a Welfare Analysis

2.1 Introduction

In recent years, consumer associations have raised concerns about the privacy of digital platforms' users as their data is increasingly collected and extensively used. These platforms often offer lower or zero prices in exchange for users' data. Personal information has become a valuable currency for firms because they can sell it to advertisers or use it to improve their offerings through product innovation. However, studies have shown that when consumers are given privacy information, they are less likely to use privacy-invasive offerings (Kummer and Schulte, 2019; Bian, Ma, and Tang, 2021; Acquisti, Brandimarte, and Loewenstein, 2015). Despite behavioral biases that may cause users to overshare data, invasive offerings come at a privacy cost.

Furthermore, different users have varying levels of privacy consciousness, and less privacy-conscious users' choices may have unintended consequences for more privacy-conscious users. This is because personal traits of the latter may be inadvertently revealed through data correlations present in the population (Acemoglu et al., 2022; Choi, Jeon, and Kim, 2019).

This paper studies the impact of this externality on welfare in a monopolist model of vertical differentiation. Specifically, the study focuses on a scenario where the platform's quality depends on the information released by users and where this information is also a source of revenue for the firm. In the presented framework, users are heterogeneous for both their willingness to pay for the service and their privacy cost, and the two are positively correlated. An element of novelty in this setting is the introduction of a negative externality, which is modeled as a network effect.

The analysis shows that when there is no externality, the monopolist under-provides privacy, and there is a downward distortion of the quality level. Moreover, when the externality is introduced, and consumers are unaware of it when they make their joining decision, it further aggravates this under-provision of privacy and increases the welfare loss. However, if they are aware and have an outside option, the monopolist may switch to the price channel by setting a zero disclosure policy, thereby eliminating the quality distortion and leaving a price mark-up driven welfare loss. Interestingly, our results suggest that introducing a negative externality may increase welfare in the market compared to the no-externality case if consumers consider its impact at the joining stage and the externality is strong enough to increase the sensitivity of demand to data disclosure over the sensitivity to prices.

Overall, our study provides insights into the complex interactions between privacy, platform quality, and consumer welfare in the presence of user data correlation. The results suggest that policymakers should consider these externalities when designing privacy regulations and, when the externality is particularly impactful, focus on raising awareness by making salient privacy notices (such as Apple Privacy Nutrition Labels) and the use of data models able to infer consumer data from minimal information.

The structure of the paper is as follows: Subsection 2.1.1 presents a short literature review, Section 2.2 introduces the basic model, and Section 2.3 finds the optimal allocation that a planner would seek when consumers are aware and when consumers are unaware. Next, section 2.4 computes the welfare of the market allocation when consumers are unaware of the externality. Then, section 2.5 extends this basic model to the case where consumer awareness is raised, and consumers internalize the externality when they make their joining decision. Finally, Section 2.6 compares total welfare in the three previous sections and explains the welfare loss, while Section 2.7 presents and discusses the results.

2.1.1 Literature

This article draws on three different areas of literature. The first focuses on the optimal quality provision in vertical differentiation models with network effects. The second examines the impact of privacy concerns on platform quality in these models. Finally, a third emerging area highlights the effects of information externalities in data markets.

The theme of the optimal quality provision in vertical differentiation is investigated in Spence (1975) that analyzes the quality decision of the monopolist and compares it to the one chosen by a benevolent planner. Spence (1975) highlighted how a welfare loss might arise from a quality set by the monopolist at a level that is distant from the one that social optimum would require. The monopolist always sets its quality

level based on the marginal consumer's willingness to pay for quality. In contrast, the benevolent planner would use the willingness to pay of the average consumer to maximize welfare. Hence, a quality distortion (defined in the literature as Spence Distortion) arises whenever the two differ. However, as reported in Tirole (1988), the sign of this distortion depends on the model, and the monopolist may over-provide or under-provide quality depending on which of the two willingness to pay is higher. Spence's result has then been extended to the case of network externalities by Lambertini and Orsini (2001) that shows how positive network effects would lead the monopolist to over-provide quality. This chapter contributes to the literature by adapting the Spence distortion to the theme of data disclosure and by discussing the effects of a negative data externality.

Secondly, Casadesus-Masanell and Hervas-Drane (2015) introduces the idea that privacy could represent a strategic element in a vertical differentiation framework, and it studies the impact of competition on the level of privacy provided. In the setting, consumers are heterogeneous for both their privacy cost and the value they assign to the service. A peculiarity of this model is that it endogeneizes consumers' decision of how much information to provide to the platform. Conversely, firms can earn revenues by selling this information in a competitive secondary market or charging consumers a price. The model results show that the level of competition positively influences the level of privacy consumers get. A related model is the one of Bloch and Demange (2018) that also treats a situation where a monopolist platform faces consumers heterogeneous on the privacy cost. Differently from Casadesus-Masanell and Hervas-Drane (2015), however, this model uses a homogeneous value for the service, and the data collection decision resides entirely on the platform. Under these assumptions, they show that the firm, for some parameter values, picks a high data exploitation level and decides to uncover the market by excluding the high-privacy-cost consumers. They also show that data collection may be excessively high from a welfare perspective. Furthermore, they expand the basic model with different policy instruments such as taxes and opt-out options. A similar comparison of policy instruments in a setting in which a monopolistic platform monetizes only disclosing personal information to third parties and can invest in quality is proposed by Lefouili and Toh (2017). They show that the monopolist always under-supplies privacy. This chapter's contribution to this literature resides in reinforcing the result related to an under-supply of privacy in a different setup and in showing that when willingness to pay for the base service is perfectly correlated to the privacy cost, the part of consumers that get excluded is the left tail of the distribution. [1]

[1] If privacy is a superior good, this assumption makes more sense than the one without correlation. However, that would be helpful to obtain a closed-form solution of the model without having to resort to simulations.

In this analysis, we include a third category of research focusing on the effects of data correlation on consumer valuation and social welfare. These studies, conducted by Acemoglu et al. (2022) and Choi, Jeon, and Kim (2019), demonstrate that data correlation can have negative consequences. Acemoglu et al. (2022) proposes a model in which a monopolist firm contracts with different users, and he can exploit the negative externality they exert on each other to minimize the price paid for consumers' information. This results in data being undervalued due to market failure. Choi, Jeon, and Kim (2019) instead analyzes a model very similar to Bloch and Demange (2018), but where consumers' base valuation of the service drives heterogeneity and, conversely, consumers are homogeneous for the impact of privacy features on utility (benefits and costs). One of the results of this paper is that the welfare loss is driven by the difference in social marginal cost and private marginal cost, which is, in turn, mainly determined by the nuisance of data collection on non-users. Consequently, the monopolist ends up over-collecting data and serving too many consumers. We contribute to this discussion by showing that the Spence distortion is aggravated by a small externality even when consumers are aware, the number of users served by the platform is optimal, and non-users are not suffering from the externality. Surprisingly, when consumers are aware, a significant externality reduces the Spence distortion and pushes the firm to offer optimal quality. Consequently, in such situations, welfare is increased by the presence of a negative externality.

Further extensions of the model may go in the direction of Bloch and Demange (2018) and Bourreau, Caillaud, and De Nijs (2018) that have analyzed taxation of a digital monopoly platform, and this theme could be applied to define the effects of an optimal Pigouvian tax and opt-out policy or a cap on data disclosure (Lefouili and Toh, 2017). [2]

2.2 Basic setup

2.2.1 Consumers

Let us consider a market where a monopolist platform faces consumers that are distributed over the support $[\bar{\theta} - 1, \bar{\theta}]$ for their taste parameter θ. This assumption is used along all variants of the model presented. Keeping fixed the support of the distribution implies that the monopolist is more likely to find it profitable to cover the market. Instead, suppose consumers' willingness to pay followed a distribution with a higher standard deviation. In that case, the monopolist may find it profitable to focus

[2]However, to introduce such elements, the model structure shall be simplified, such as in Bloch and Demange (2018).

only on the distribution's right tail. [3]

Consumers value the platform quality based on the information exchanged (or activity) (y) and on the rate of information disclosure (d) to advertisers, level of which is set by the platform. If we indicate with Y the total stock of information collected by the firm, we assume that consumers' utility function is:

$$U_i = \begin{cases} \theta(y - y^2 - yd) - \alpha Y d - P & \text{if the consumer buys,} \\ -\beta\alpha Y_{-i} d & \text{if the consumer does not buy,} \end{cases} \tag{2.1}$$

Equation 2.1 illustrates the consumer's utility in case they choose to buy the product or not.[4] In this equation, P denotes the good's price (or the one-time membership fee), while d represents the disclosure rate of the consumer's provided information y. Furthermore, the parameter α can be interpreted either as the intensity of user data correlation among consumers or as the weight of the disutility from the externality on users, while the parameter β has a similar interpretation for non-users.

With $\alpha = \beta = 0$, this model would be the standard setup in Casadesus-Masanell and Hervas-Drane (2015), where consumers are heterogeneous for their taste parameter that expresses their willingness to pay for the good provided and for privacy. A peculiarity of this utility function is that consumers with a higher willingness to pay for the good also have a higher distaste for disclosure. This correlation assumption between the two components is supported if privacy is a superior good.[5] Finally, this utility function is concave in y and depends negatively on d and P that the firm sets.

A difference from the Casadesus-Masanell and Hervas-Drane (2015) framework is the presence in the utility function of a negative externality that arises from the fraction of total information sold by the firms in the ads market ($Y \cdot d$). This term is introduced as the multiplication of the whole information stock accumulated by the firm (Y) by the disclosure rate applied (d). The total stock of information Y is given by the product of the information provided by the single user y and the demand x for the good.[6] This $Y \cdot d$ term is then weighted by a parameter α, which can be seen either as the strength of data correlation in the population or as the weight of this externality on utility.

As in Choi, Jeon, and Kim (2019), also non-users are negatively impacted by the

[3] One additional implication worth noticing is that with the price of data standardized to one, we lose the possibility to vary the relative value of the two revenue sources that are employed by the monopolist (see 2.2.2).

[4] We use the terms buy and join interchangeably throughout the text. This is because the price can be seen as a one-time membership fee to subscribe to the platform or the price paid for a good.

[5] However, there could be cases where a consumer that gets more utility from the platform also prefers more disclosure of his data. In this case, the model of Bloch and Demange (2018) would be informative of the equilibrium

[6] y could be seen as the user's activity level on the platform, such as the number of likes, comments, and posts.

externality $\alpha Y_{-i}d$ that is weighted by $\beta \in [0, 1]$. Since, in this case, the consumer does not join, Y_{-i} is the total information stock provided by all users, but i: $Y_{-i} = yx_{-i}$ where x_{-i} represents the demand of all other users except user i, and y is the level of information revealed by the user.

2.2.2 Firm's profit, stages, and equilibrium concept

As in Casadesus-Masanell and Hervas-Drane (2015), the monopolist has two revenue sources: prices and data. Its profit function takes the form:

$$\pi = x(P + yd), \tag{2.2}$$

where x is the demand for the service, P the price paid, d the disclosure rate and y is the consumer's provided information.

The game is a classic location-then-price one, with the addition of the information stage. We propose two variations: one with consumers aware of the externality when they join the platform and another where consumers are unaware of the externality at all stages. When consumers are unaware, the stages of the game are:

- At t=0, the firm sets its quality level by deciding on the disclosure d
- At t=1, the firm sets its membership price P
- At t=2, consumers observe P, d and decide whether or not to join
- At t=3, consumers set the level of information y

In this first variant of the game, consumers are unaware at $t = 2$ of the joining decision of other agents and do not account for the externality parameter when taking their joining decision. Only after $t = 3$ the actual utility level (externality included) is observed.

Alternatively, when consumers are aware of the externality, the game stages are:

- At t=0, the firm sets its quality level by deciding on the disclosure d
- At t=1, the firm sets its membership price P
- At t=2, consumers observe P, d and decide whether or not to join, also considering the decision of others (the demand enters the indifferent consumer as in Lambertini and Orsini (2001))
- At t=3, consumers set the level of information y

In this second variant, instead, consumers are forward-looking and perfectly anticipate the joining decision of other users. They are also perfectly aware of the externality effect and consider this factor when taking the decision at $t = 2$.

We look for the Sub-Game Perfect Nash Equilibrium by backward induction and the equilibrium defined as the triplet: price, disclosure, and information released by consumers (P^*, d^*, y^*).

Before turning to the solution of the two monopolist games just described, we look into the optimal allocation of a benevolent planner that sets P and d to maximize social welfare.

2.3 First Best

In this Section, we highlight what a perfectly informed, benevolent government would choose when it keeps into account the externality.

Welfare follows the traditional sum of consumer surplus and profits to which we need to add the "non-consumers" negative surplus derived from the externality. Hence, we have the following:

$$\max_{d,P} \left\{ W^c = \int_{\bar{\theta}-1}^{\bar{\theta}} U_i^u \, d\theta + \pi \right\}, \tag{2.3}$$

that when $\theta^* \in (\bar{\theta} - 1, \bar{\theta})$ and the non-users disutility $U_i^n \leq 0$ enters the welfare maximization problem becomes:

$$\max_{d,P} \left\{ W^u = \int_{\theta^*}^{\bar{\theta}} U_i^u \, d\theta + \int_{\bar{\theta}-1}^{\theta^*} U_i^n \, d\theta + \pi \right\}, \tag{2.4}$$

depending on how consumers react to the externality (i.e., whether they are aware or not), the definition of the indifferent consumer changes.

2.3.1 Information stage and indifferent consumer

In the spirit of Acemoglu et al. (2022) we build the models solved in this essay with the following assumption: the single user is infinitesimal to the full support of the distribution.

Therefore, its own demand is negligible with respect to the whole demand and the value of a single user's information and joining decision is irrelevant to the total information stock. Because of this assumption we can approximate the stock of information in the following way:

$$Y_{-i} = y^* x_{-i} \approx y^* x = Y \tag{2.5}$$

where y^* is the equilibrium level of information as determined in the remaining of this Section. This assumption is rooted on the reasoning of Acemoglu et al. (2022) that states: *"[...] the marginal increase in the leaked information from individual i's sharing decision is decreasing in the information shared by others. This too is intuitive and follows from the fact that when others' actions reveal more information, there is less to be revealed by the sharing decision of any given individual."*.

The assumption finds also support in the geographical widened coverage of plat-

forms and the presence of data brokers. These two components strongly support the idea that a single consumer's information is insignificant when compared to the overall stock of information. [7]

Given that with both consumer awareness and unawareness, this assumption implies that the externality is not internalized at the information level, consumers' equilibrium level of information is:

$$y^* = \frac{1-d}{2}, \tag{2.6}$$

inserting this into the utility function, we have

$$U = \theta\frac{(1-d)^2}{4} - P - \alpha Y d,$$

given that the joining decision is different in the two variations, we analyze them separately.

Unaware Consumers Firstly, when consumers are unaware, the externality effect on utility will be revealed only when they use the service and observe the perfect targeting of ads. Therefore, in monopoly, the indifferent consumer remains similar to Casadesus-Masanell and Hervas-Drane (2015) case:

$$\theta^* = \frac{4P}{(1-d)^2}, \tag{2.7}$$

Aware Consumers When consumers are aware of the consumer that is indifferent between buying and not buying solves the following:

$$\begin{aligned} U_{buys} &\geq U_{not\ buy}, \\ \theta\frac{(1-d)^2}{4} - P - \alpha Y d &\geq -\beta\alpha Y d, \\ \theta\frac{(1-d)^2}{4} - P &\geq \alpha x\frac{1-d}{2}d(1-\beta), \end{aligned}$$

where the last equation uses $Y = xy^*$.

In this case, we find it helpful to express the model in terms of $e = \alpha(1-\beta)$: given that $1-\beta$ determines the "saving" of utility that the consumer can get by not joining (the consumers' outside option), this term multiplied by α represents the *net impact of the externality on demand*. At the extreme, if the externality is unavoidable by the consumer $1-\beta = 0$, then the net impact of the externality on demand is zero.[8]. Thus, we use e to simplify the equations.

[7] Additionally, it provides modeling benefits of simplifying the modeling process, such as the simplification of the indifferent consumer equation, that outweigh the need for strict mathematical rigor.

[8] This is a crucial part of the elasticity of the demand function to the externality parameter

To identify demand, we employ the straightforward approach of Lambertini and Orsini (2001), that is, we consider x as the integral over the space $[\hat{\theta}, \bar{\theta}]$ of the density function and we substitute this in the indifferent consumer equation.

The latter becomes:

$$\theta \frac{(1-d)^2}{4} - P - e\frac{1-d}{2}d\int_{\hat{\theta}}^{\bar{\theta}} f_\theta d\theta = 0, \tag{2.8}$$

with $\hat{\theta} = max[\bar{\theta} - 1, \theta^*]$, we have two cases: $\hat{\theta} = \bar{\theta} - 1$ that is full market coverage (FMC), and partial market coverage (PMC) that corresponds to the case in which $\hat{\theta} = \theta^* > \bar{\theta} - 1$.

In the FMC case, the indifferent consumer writes:

$$\theta^*_{FMC} = \frac{4\left(P + e\frac{1-d}{2}d\right)}{(1-d)^2}, \tag{2.9}$$

while in the PMC case, the equation becomes the following:

$$\theta^*_{PMC} = \frac{P + e\bar{\theta}\frac{1-d}{2}d}{\frac{(1-d)^2}{4} + \frac{(1-d)}{2}ed}, \tag{2.10}$$

The indifferent consumer in the unaware case is as defined in (2.7) and is defined by solving for θ equation (2.8) when consumers are aware, obtaining the conditions (2.9) and (2.10). Despite this change, the only difference between the aware and unaware first bests is that the aware consumers model first best requires a lower price. This is because aware consumers have more bargaining power and demand higher compensation for their data. Nonetheless, P is only a transfer, and this difference does not impact final welfare. Therefore, we relegate to the Appendix the planner's optimization problem for the 'aware' case (see Section B.4).

Additionally, while with a covered market, we can find a closed-form solution for the (P, d) pair to the maximization problem, the uncovered market maximization problems (both unaware and aware) become a tedious task. In these cases, we need to rely on numerical simulations through Python and Mathematica (see Section B.3 and Section B.4).

By studying the numerical simulations tables reported in the Appendix, it results that to maximize welfare, the planner will always rely on a covered market configuration as the uncovered market simulation of welfare tends to the covered one in both cases. Therefore, a covered solution always gives more welfare than an uncovered one,

whatever α and β values are. [9] [10]

2.3.2 Covered market with unaware consumers

In this case, the externality consumers suffer does not impact the market demand, and since the market is covered, θ^* will only influence the price.

When the market is covered, the welfare function writes:

$$W_c = \int_{\bar{\theta}-1}^{\bar{\theta}} \left(\theta \frac{(1-d)^2}{4} - \alpha d \frac{1-d}{2} - P \right) d\theta + \pi,$$

and by solving the integral, we obtain the following:

$$W_c = \frac{1}{8}(d-1)(d(4\alpha + 2\bar{\theta} - 5) - 2\bar{\theta} + 1), \tag{2.11}$$

by studying this polynomial, we notice that when the market is covered, welfare is price-neutral ($\frac{\partial W_c}{\partial P} = 0$), and prices are only a transfer of surpluses. Thus, the planner picks the price that keeps the market covered and then maximizes welfare through the disclosure rate.[11]

By maximizing this function in d, we obtain the FOC:

$$\frac{1}{4}(\alpha(4d-2) + d(2\bar{\theta} - 5) - 2\bar{\theta} + 3) = 0,$$

with the following SOC:

$$\frac{\partial^2 W}{\partial d^2} = \alpha + \frac{\bar{\theta}}{2} - \frac{5}{4} \leq 0 \iff \alpha \leq \frac{5}{4} - \frac{\bar{\theta}}{2}$$

that is always satisfied as:

$$0 \leq d_c^w \leq 1 \iff \alpha \leq \frac{3}{2} - \bar{\theta},$$

given the bounds on disclosure (0,1), we would need an α larger than $\frac{3}{2} - \bar{\theta}$ to violate the second order condition. However, when α is larger than $\frac{3}{2} - \bar{\theta}$, the optimal disclosure becomes negative, so that welfare is maximized by $d = 0$, and we can be sure that the SOC is satisfied. Therefore, we can state the following Proposition that collects the

[9] We defer to a later stage the analytical proof of this result.

[10] The only exception found is in the model with aware consumers around the point $(\alpha, \beta, \theta) = (1, 0.4, 1.2)$ where a point of discontinuity arises, and it causes an uncovered welfare solution

[11] Notice that this would not hold if the distribution support were not fixed. In that case, welfare is not price-neutral, and that $\frac{\partial W_c}{\partial P} > 0$ with $\frac{\partial^2 W_c}{\partial P^2} \geq 0$ hence the planner would take the highest possible price to keep the market covered.

results of welfare maximization when the market is covered:

Proposition 3. *The welfare maximizing planner would set a price that ensures a covered market:*

$$P_{fb} = \frac{1}{4}(\bar{\theta} - 1)(1 - d)^2$$

and then uses disclosure to maximize welfare according to the rule:

$$d_{fb} = \begin{cases} \dfrac{2(\alpha + \bar{\theta}) - 3}{2(2\alpha + \bar{\theta}) - 5} & \text{if } \ 1 \leq \bar{\theta} \leq \dfrac{3}{2} \wedge \alpha \leq \dfrac{1}{2}(3 - 2\bar{\theta}), \\ 0 & \text{Otherwise,} \end{cases}$$

this gives welfare as a function of α and $\bar{\theta}$:

$$W_{fb} = \begin{cases} \dfrac{(\alpha - 1)^2}{10 - 8\alpha - 4\bar{\theta}} & \text{if } \ 1 \leq \bar{\theta} \leq \dfrac{3}{2} \wedge \alpha \leq \dfrac{1}{2}(3 - 2\bar{\theta}), \\ \dfrac{1}{8}(2\bar{\theta} - 1) & \text{Otherwise,} \end{cases} \tag{2.12}$$

further inspection of the denominator of the welfare function in this interval shows that $W_c > 0$.

The proof follows from the maximization in d of the function in (2.11), and from the comparison with the results of numerical simulations to show that the market is always covered (see Section B.3). [12]

By analyzing the inequality constraint in the optimal d rule, the welfare-maximizing disclosure rate can be positive only if it respects two necessary conditions. Given that $\alpha > 0$, the first necessary condition to have a positive disclosure rate is that the wealthiest consumer is poor enough, i.e., $\bar{\theta} \leq \frac{3}{2}$. Additionally, given that the minimum acceptable value for the willingness to pay of the wealthiest consumer is $\bar{\theta} = 1$, and anything below this value is not acceptable (since $\underline{\theta} = \bar{\theta} - 1 \geq 0$): the externality cannot be larger than $\frac{1}{2}$.

Corollary 1. *When the market is covered, a positive disclosure rate maximizes welfare only in "poor markets" (where the willingness to pay of the richest consumer is not too high $\bar{\theta} \leq \frac{3}{2}$), and in those markets where the externality is less significant: $\alpha < \frac{1}{2}$. In all other cases, the welfare-maximizing disclosure rate shall be zero.*

[12] It is, however, likely that this optimal configuration is not unique as we used P to keep the market covered while the planner may use d to achieve this.

2.4 Market allocation and welfare with unaware consumers

In this Section, we study the case of unaware consumers. In this case, consumers' equilibrium level of information y^* is described by (2.6), and the indifferent consumer is defined by eq (2.7). Consequently, given that the externality does not impact these two equations, the monopoly solution computed in Casadesus-Masanell and Hervas-Drane (2015) paper remains unaffected.

We now characterize the total welfare of this solution while we delay to Section 2.5 the discussion of the case where consumers are aware.

2.4.1 Market allocation and welfare with unaware consumers

Here we analyze the welfare properties of the market allocation of the basic model.

Covered market welfare analysis

Given that the indifferent consumer and the profit function are the same, the market allocation does not change. We can therefore state the following remark:

Remark 1. *If the market is covered, and consumers are distributed according to a uniform distribution over the interval* $[\bar{\theta}-1, \bar{\theta}]$*, we have two cases:*

$\bar{\theta} \geq 2$ *then monopolist's optimal disclosure is* $d_c^m = 0$ *and price* $P_c^m = \dfrac{\bar{\theta}-1}{4}$*. Profits become* $\pi_c^m = \dfrac{\bar{\theta}-1}{4}$*, and the market generates a consumer surplus of* $CS_c^m = \dfrac{1}{8}$ *and total welfare* $W_c^m = \dfrac{1}{8}(2\bar{\theta}-1)$,

$1 \leq \bar{\theta} \leq 2$ *then monopolist's optimal disclosure rate is* $d_c^m = \dfrac{\bar{\theta}-2}{\bar{\theta}-3}$*, with a price of* $P_c^m = \dfrac{\bar{\theta}-1}{4(\bar{\theta}-3)^2}$ *and profits become* $\pi_c^m = \dfrac{1}{12-4\bar{\theta}}$*, the market generates a consumer surplus of* $CS_c^m = \dfrac{4\alpha(\bar{\theta}-2)+1}{8(\bar{\theta}-3)^2}$*, and total welfare* $W_c^m = \dfrac{4\alpha(\bar{\theta}-2)-2\bar{\theta}+7}{8(\bar{\theta}-3)^2}$,

the proof follows directly from the proof of Proposition 1 in Casadesus-Masanell and Hervas-Drane (2015), to which we applied the definition of welfare as:

$$W^c = \int_{\bar{\theta}-1}^{\bar{\theta}} U_i^u \, d\theta + \pi, \tag{2.13}$$

If $\bar{\theta} < 2$, the externality negatively affects welfare as expected, whereas when $\bar{\theta} \geq 2$ optimal disclosure is zero, and the price is the only source of revenues, and neither consumer surplus nor total welfare is affected by the value of the externality so we do not focus on this parameter region for the remaining of the article.

Uncovered Market - Welfare Analysis

Given the indifferent consumer in (2.7) and optimal information choice, the profit function reduces to:

$$\pi_m^u = \left(\bar{\theta} - \frac{4P}{(d-1)^2}\right)\left(P - \frac{(d-1)}{2}d\right),$$

by maximizing in P this function, we find that $P = \frac{1}{8}(d-1)(\bar{\theta}d - \bar{\theta} + 2d)$, and inserting it into the profit function gives us the optimization problem:

$$\max_d \frac{1}{16}(\bar{\theta} + d(2-\bar{\theta}))^2$$

the maximization of this function shows that a solution exists only when $\bar{\theta} \geq 2$ and when the support of the distribution is smaller than one because otherwise, the profit function is convex in d and the only candidate equilibrium would be full disclosure ($d = 1$). Indeed, a so high disclosure rate would cause consumers not to use the platform at all: when $d = 1$, every consumer's optimal level of activity is zero ($y^* = 0$). Therefore, full disclosure ($d = 1$) completely degrades quality and destroys the potential profit. As a consequence, the location of the indifferent consumer tends to infinity when quality is so low: $\frac{4P}{(1-d)^2} \to \infty$.

Finally, we can state the following remark:

Remark 2. *When the support of the distribution is fixed at one, there is not a feasible solution to the monopolist problem that involves an uncovered market configuration.*

The proof was omitted as it directly follows from the proof of Proposition 1 in Casadesus-Masanell and Hervas-Drane (2015) and the assumption about the externality. Because of Remark 2, we can generalize the subscripts of the welfare function from Remark 1 to be $W_c^m = W_m$.

As in Casadesus-Masanell and Hervas-Drane (2015), poorer markets are those markets where the mean willingness to pay for the services is lower: an example of this would be the search engine markets where the standard price in the industry is zero against a music streaming platform where consumers' higher willingness to pay brings in the market ad-based offerings (YouTube) together with those subscription based (Spotify).

An alternative interpretation, which opens possibilities for empirical estimation, is considering countries as different markets. In this case, we expect the monopolist platform to have high disclosure rates in low-income countries and low disclosure rates in high-income countries. This may explain the cross-national differences in privacy policies observed in Kumar et al. (2022).

2.5 Market allocation and welfare with aware consumers

In this Section, the utility and profit structure is the same as the monopolist solution provided in Casadesus-Masanell and Hervas-Drane (2015) framework. However, we add informational externalities on the consumers' side and explore the possibility that perfectly rational consumers, having observed disclosure rates and prices, can anticipate the information stock that the firm will accumulate flawlessly. They can "internalize the externality" at the game's third stage. In other words, the effect arises when consumers release information at the information stage (last stage) and do not consider their impact on others' utility.

However, when consumers decide whether to buy or not, they perfectly anticipate the information stock at the firm disposal. They are aware of the impact of the externality, so they change their consumption accordingly. Therefore, this "negative network effect" is introduced similarly to Lambertini and Orsini (2001) and is partially internalized.[13]

The information stage remains as in Section 2.3.1, and the indifferent consumer when the market is fully covered is described by Eq (2.9) and Eq (2.10).

In both cases, the monopolist has the following profit function:

$$\pi = x(P + \frac{1-d}{2}d),$$

this is the profit function of Casadesus-Masanell and Hervas-Drane (2015) whose disclosure revenues satisfy the concavity conditions of Choi, Jeon, and Kim (2019).

2.5.1 Full market coverage case

In the case of full market coverage $\hat{\theta} \leq \bar{\theta} - 1$ and consequently $x_{fmc} = 1$. This simplifies the problem, and we can find the analytical solution that we state in Lemma 1:

Lemma 1. *When consumers are aware of the externality and the market is covered, the monopolist charges:*

$$P^*_{fmc} = \begin{cases} \dfrac{(1-e)\left(2e^2 - e(\bar{\theta}-3) + \bar{\theta} - 1\right)}{4(2e+\bar{\theta}-3)^2} & \text{if } \ \bar{\theta} + e \leq 2 \ \wedge \ \bar{\theta} \neq 3 - 2e, \\ \dfrac{\bar{\theta}-1}{4} & \text{Otherwise,} \end{cases}$$

[13] An alternative would be to use the responsive rational expectation assumption, which would work similarly and achieve the same results.

and disclosure rate:

$$d^*_{fmc} \begin{cases} \dfrac{e+\bar{\theta}-2}{\bar{\theta}+2e-3} & \textit{if} \quad \bar{\theta}+e \leq 2 \ \wedge \ \bar{\theta} \neq 3-2e, \\ 0 & \textit{Otherwise,} \end{cases}$$

optimal profit is:

$$\pi^*_{fmc} = \begin{cases} \dfrac{(1-e)^2}{4(3-\bar{\theta})-8e} & \textit{if} \quad \bar{\theta}+e \leq 2 \ \wedge \ \bar{\theta} \neq 3-2e \\ \dfrac{\bar{\theta}-1}{4} & \textit{Otherwise} \end{cases}$$

where $e = \alpha(1-\beta)$ is the net impact of the externality on demand.

the proof is the standard backward induction procedure to solve the game and is presented in Section B.1.

When the market is covered and consumers are aware, A positive disclosure rate can emerge only under two conditions. Firstly, even when e is at its minimum, the market must be "poor", i.e., willingness to pay of richest consumer $\bar{\theta} \in (1,2)$. Given a uniform distribution with fixed support, this condition implies a low mean willingness to pay in the market. The consumer with the highest willingness to pay is relevant only as a proxy for the mean of the distribution. That is what defines a market as rich or poor: so we can directly link the w.t.p. of the richest consumer to the market value for the monopolist. Secondly, the net effect of the externality on demand cannot be too large relative to the mean willingness to pay (market value). From the constraint $\bar{\theta}+e \leq 2$, we see that as e reaches its upper bound of one, the willingness to pay that can sustain a positive disclosure and covered market becomes smaller and smaller.

By decomposing e into its components, we see that one effect of this model, where consumers are aware of the externality, is how β impacts this constraint: when β is low, the constraint on α gets tighter. Thus, the better the outside option for users (higher $1-\beta$), the more likely the optimal disclosure rate is zero.

Not surprisingly, then, profit is decreasing in the value of externality among the users ($\frac{\partial \pi}{\partial \alpha} < 0$) and increasing in the value of externality among the non-users ($\frac{\partial \pi}{\partial \beta} > 0$) because the more substantial the externality among non-users is, the worse the outside option for potential subscribers. At the extreme $\beta = 1$, consumers have no outside option and can only bear the externality.

A practical example of how the externality effect may propagate to non-users may be the cases of Facebook and WhatsApp, or Gmail and Google Search. In both cases, one of the products that the multiproduct firm offers is a kind of an essential facility for most of the consumers in the digital economy, and if one has a particular high evaluation of privacy may decide not to use Gmail but still finds the value of Google

Search reduced by the externality other users cause. Therefore, one also ends up joining Gmail because of the powerlessness in front of the external effect. This is a similar mechanism to Acemoglu et al. (2022), where the platform extracted more data from consumers by leveraging this externality among consumers in contracting with the platform.

When the market is covered, the effect on prices and disclosure of the externality parameter (α) is mediated by the users' outside option and by their willingness to pay for quality (β and $\bar{\theta}$). We show the effect of different values of these two parameters on optimal prices and disclosure in a covered market in Figures 2.1a and 2.1b. These two figures show how the average willingness to pay in the market shifts the monopolist's incentive to use disclosure versus prices as the primary revenue instrument. As the average willingness to pay is perfectly correlated with the consumers' distaste for disclosure, absent the externality, the monopolist uses prices in markets with a high willingness to pay for quality and disclosure in markets where this is low. When the market is richer, consumers are more reactive to disclosure than to prices, and the monopolist extracts more revenues with the monetary instrument than the data instrument. On the contrary, in very poor markets, as the sensitivity of consumers to the disclosure parameter is lower than the price sensitivity, the monopolist favors disclosure.

When we introduce the externality parameter, it shifts the indifferent consumer to the right. Consequently, given the interaction of the disclosure with the externality the sensitivity of consumers to disclosure rises, for the monopolist disclosure revenues become more expensive in terms of extensive margin. In relatively wealthier markets, this effect quickly drives the disclosure rate to zero (orange and red curves) and causes the monopolist to increase prices further.

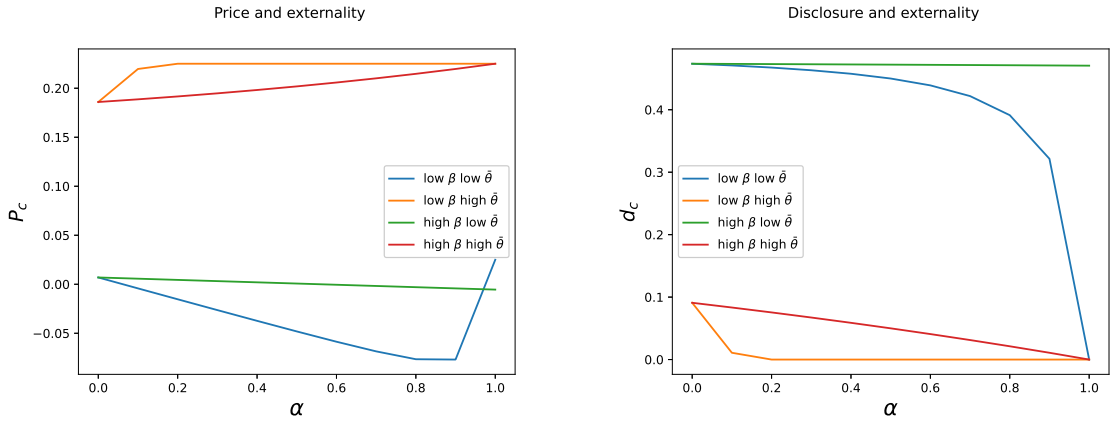
However, we observe the most significant impact of the externality in markets where a positive (and high) disclosure rate was set before introducing the externality. In poor markets, in fact, the drop in disclosure rate and the correspondent switch to prices tend to be quite dramatic (blue line). With the externality, prices start to drop in poor markets and reach the negative area (consumer subsidization).[14] This is because in these markets, the monopolist that faces aware consumers would need to compensate them for the externality through negative prices to keep extracting revenues in the ads market through data sales. Then, when the distaste for disclosure dominates prices, the loss in extensive margin from further raising disclosure dominates the one that the monopolist would have from higher prices. Therefore, at this point, the monopolist switches revenue sources.

This switch happens at a rate directly proportional to the consumers' outside op-

[14]This compensation can be a non-monetary transfer from the firm to consumers: it can be seen as services provided to consumers below costs or coupons ad discounts.

tion. If consumers do not have a good alternative option, the externality impact on demand will be less (shown by the green and red lines). This means that both price and disclosure will be less affected by large values of the externality, resulting in smoother and less noticeable changes.

Figure 2.1: Prices, disclosure and α relationships for different values of $\beta = \{1/10, 9/10\}$ and $\bar{\theta} = \{11/10, 19/10\}$



(a) Price and α (b) Disclosure rate and α

2.5.2 Partial market coverage

While the function π^a_{fmc} in Lemma 1 represents the profit function when consumers are aware and the market is covered, we still need to compare it with the PMC case. A comparison that we present in this subsection.

Considering the partial market coverage case with $\hat{\theta} = \theta^*$, the indifferent consumer is defined from eq.(2.10). By inspecting this condition, we observe that an uncovered market is possible in this model even when $P < 0$, which was not feasible in the unaware model.

It is then straightforward to obtain demand and the profit function:

$$x^u = \frac{(d-1)^2\bar{\theta} - 4P}{(d-1)(d(1-2e)-1)}, \tag{2.14}$$

$$\pi_u = \frac{(d^2 - d - 2P)\left((d-1)^2\bar{\theta} - 4P\right)}{2(1-d)(d(1-2e)-1)}, \tag{2.15}$$

whose maximization in price results in the following optimal price:

$$P^a_u = \begin{cases} \frac{1}{8}(d-1)(d(\bar{\theta}+2) - \bar{\theta}) & \forall(\theta, e, d) \in \mathcal{R}, \\ \text{Indeterminate} & \text{Otherwise}, \end{cases} \tag{2.16}$$

where the region $\mathcal{R}$ is a parametric region that was obtained by imposing the uncovered

market condition ($\bar{\theta} - 1 < \theta^* \leq \bar{\theta}$) on the profit-maximizing price.

This region is defined as the set of points (θ, e, d) satisfying either one of the following conditions:

- Condition 1: $\begin{cases} \theta \geq 2 \\ \frac{1}{2} < e \leq 1 \\ \hat{d} < d < 1 \end{cases}$
- Condition 2: $\begin{cases} 1 \leq \theta < 2 \\ 0 \leq e < \frac{1}{2} \\ 0 \leq d < \hat{d} \end{cases}$
- Condition 3: $\begin{cases} 1 \leq \theta \leq 2 \\ \frac{1}{2} \leq e \leq 1 \\ 0 \leq d < 1 \end{cases}$

where $\hat{d}(e, \bar{\theta}) = \frac{\bar{\theta}-2}{4e+\bar{\theta}-4}$.

Condition 1 regards rich markets: the externality impact must be large enough, and the disclosure rate must respect a lower bound to have an uncovered equilibrium in such markets. The lower bound on disclosure is a function that depends negatively on the externality value and positively on the market value.

Condition 2 and condition 3 give conditions on the acceptable disclosure range in poor markets when the externality value is low or high, respectively. In these markets, to have a result that respects the uncovered market constraint when the externality is low, the disclosure must be lower than the upper bound, which has the same properties as the lower bound presented in Condition 1. Alternatively, when the externality is high, the disclosure rate is unbounded and can be between zero and one.

When the parameter values fall in $\mathcal{R}^{\complement}$, the openness of the set, implied by the price inequality in the uncovered market condition, would grant no feasible equilibrium, and the price maximization problem has no solutions.

Finally, the profit function and the maximization problem at the disclosure stage are:

$$\pi_u^a = \begin{cases} \dfrac{(1-d)(\bar{\theta} - d(\bar{\theta}-2))^2}{16(d(2e-1)+1)} & \forall (\theta, e, d) \in \mathcal{R}, \\ \text{Indeterminate} & \text{Otherwise}, \end{cases}$$

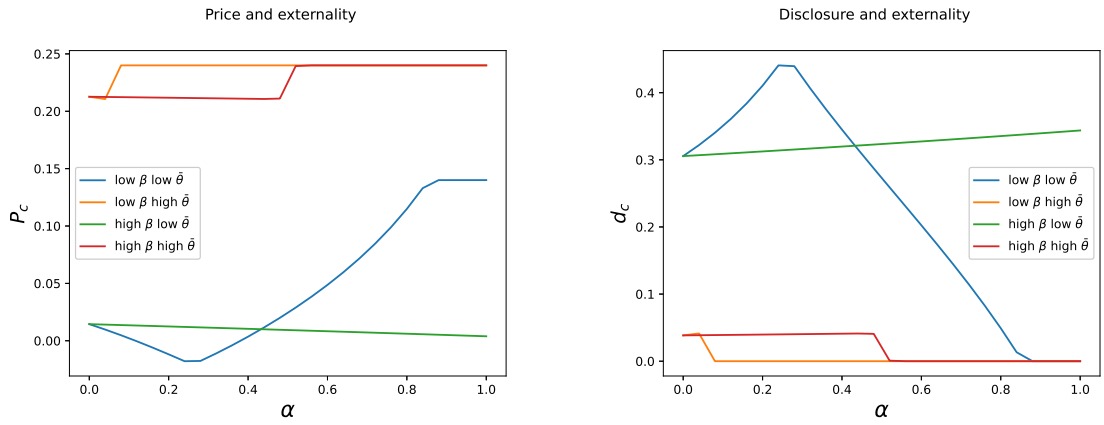
$$\begin{array}{llll} \max_d & \pi_u^a & \text{s.t.} & d \leq 1, \\ & & & d \geq 0, \end{array}$$

we solve the problem numerically through scipy *optimize* package with Sequential Least Square programming and through *"NMaximize"* command in Wolfram Mathematica with *"SimulatedAnnealing"* and *"DifferentialEvolution"* methods.[15] The analytical solution to this problem becomes algebraically irksome and would only clutter the text without offering additional insights compared to the simulation results. Therefore, we only explain the characteristics of the equilibrium without providing an analytical solution.

The numerical simulation with an uncovered market configuration and aware consumers shows that for low values of the net externality e and rich markets $\bar{\theta} \geq 2$, the candidate optimal price would violate the lower bound of $\bar{\theta} - 1 < \theta^* \leq \bar{\theta}$. Similarly, to Casadesus-Masanell and Hervas-Drane (2015), the openness of that set implies that no equilibrium exists.

Figure 2.2a and 2.2b help visualize the simulation results and the profit-maximizing rule followed by the monopolist in different types of markets. The most significant difference is undoubtedly in poor markets with an outside option: in this case, α has a non-monotonic effect on disclosure (and price) that first increases (decreases) in α (for $\alpha \leq 0.3$) and then drops (increases). Noticeably, with an uncovered market configuration, the prices never go into the negative area, and the monopolist does not subsidize consumers. In sufficiently rich markets ($\frac{3}{2} \leq \bar{\theta} < 2$), instead, when the externality is strong $e > \frac{1}{2}$, the monopolist optimally sets zero disclosure and positive prices under an uncovered market configuration.

Figure 2.2: Prices, disclosure and α relationships for different values of $\beta = \{1/10, 9/10\}$ and $\bar{\theta} = \{11/10, 19/10\}$ with an uncovered market



(a) Price and α

(b) Disclosure rate and α

We now proceed to specify the optimal decision of the monopolist when he chooses between a covered and uncovered market.

[15] Alternatively, we could simplify the problem by assuming a standard uniform and solving the KKT to obtain a closed-form solution, but in doing so, we could no longer appreciate the differences among different markets.

2.5.3 Comparison covered-uncovered market

From the simulation results in Appendix B.2, we compare the profit obtained in a covered market (as stated in Lemma 1) and the profit obtained in an uncovered market configuration. From this comparison, we can state the following Lemma:

Lemma 2. *The decision to cover the market depends on the externality effect on the demand (e) and the willingness to pay of the richest consumer ($\bar{\theta} \geq 2$). This relationship is summarized as follows:*

1. *When $\bar{\theta} \geq 2$:*

   (a) *When e is sufficiently low ($e \lessapprox \frac{2}{10}$), the monopolist always covers the market.*

   (b) *Larger values of e and $\bar{\theta}$ make the uncovered configuration an equilibrium more frequently. Eventually, if $e \gtrapprox \frac{4}{10}$, an equilibrium with an uncovered market configuration may arise even in very poor markets.*

2. *When $\bar{\theta} \geq 2$:*

   (a) *The monopolist always covers the market.*

A closed-form analytical solution is necessary to understand the intuition behind these results quickly. The most likely explanation derives from how the externality is modeled and impacts the indifferent consumer. From Eq (2.9) and (2.10), it emerges that the higher the impact of the externality on the utility of users and the larger the shift of the indifferent consumer to the right. Consequently, the only instrument the monopolist has to counteract this effect is to reduce the disclosure rate and use only prices to earn revenues. The simulation table in the Appendix confirms this and shows that higher levels of the net externality impact on demand correspond to a reduction of the disclosure rate. Thus, also in this case, for higher values of the externality, which reduces consumers' utility proportionally to d, consumers' sensitivity to disclosure increases. Furthermore, this shift is amplified by the interaction between the externality and positive levels of disclosure that further shift the indifferent consumer's location on the right. An increase in willingness to pay causes a similar effect on the indifferent consumer: when the market is rich, consumers' evaluation of the service is higher as well as the positively correlated distaste of disclosure and the indifferent consumer shifts to the right.

In the next Section, we compare the welfare obtained by the monopolistic solutions with the first best and characterize the welfare loss.

2.6 Welfare comparison

In this section we compare the three relevant welfare function W_{fb}, W_a, W_m as stated respectively in Proposition 3, Proposition 1 and Remark 1.

In Figure 2.3 we plot the three welfare functions against the α parameter for the combination of market value ($\bar{\theta} = \frac{6}{5} \vee \frac{19}{10}$) and externality on non-users ($\beta = \frac{1}{10} \vee \frac{9}{10}$). While we use unique colors to indicate that the market is always covered for the basic unaware solution (blue) and the first best solution (green), we plot the aware case in different colors depending if it involves a covered market (orange) or an uncovered one (red). The discontinuities between the orange and the red lines are where the monopolist switches from the covered to the uncovered market configuration.

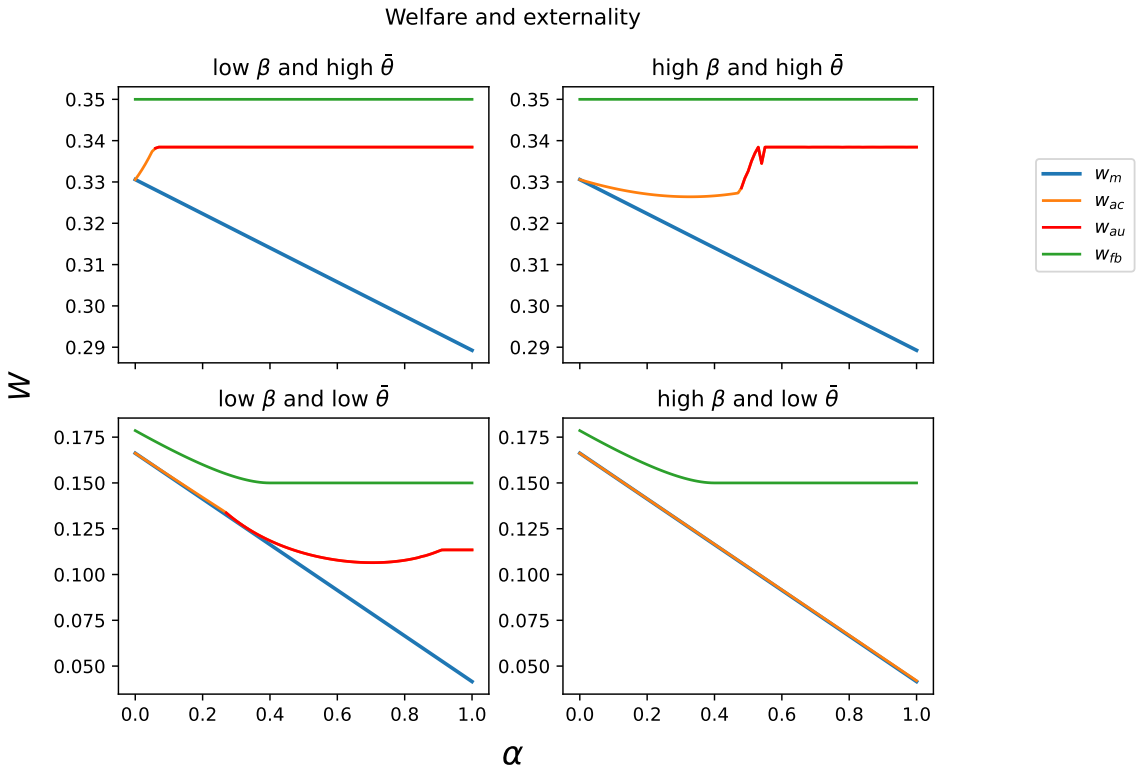
In Choi, Jeon, and Kim (2019), the welfare loss arises from the externality effect on non-users that creates a divergence between private cost and social cost, and the monopolist ends up serving too many consumers compared to the social optimum. The β high case represented in the two plots on the right of Figure 2.3 shows that also in the models presented, the absence of a viable outside option gives the bargaining power to the monopolist and results in a welfare loss. This also happens in markets where the externality is internalized (aware case). However, this is not the only source of divergence from the first best, as the left panels of this figure show. Indeed, even if the externality parameter for non-users is very low or zero, a welfare loss always arises in the unaware case, even when the consumers have a viable outside option and the monopolist serves the whole market.[16]

In Section 2.3, we have shown that the first best solution requires high prices and zero disclosure in richer markets $\bar{\theta} > \frac{3}{2}$ and lower prices and higher disclosure in poorer markets $1 \leq \bar{\theta} \leq \frac{3}{2}$. These two situations are respectively represented in the two top (rich) and bottom (poor) panels of Figures 2.4 and 2.5. Unsurprisingly, in the unaware case, the monopolist distorts prices and disclosure rates with respect to the first best solution and ends up under-supplying privacy. Indeed, while Figure 2.4 shows that the price (blue) is always too low in the unaware case with respect to the welfare-maximizing price (green), Figure 2.5 shows that the disclosure rate is always too high (blue) with respect to the welfare-maximizing one (green). This result is independent of the externality.

Therefore, differently from Choi, Jeon, and Kim (2019), the monopolist that faces unaware consumers always serves the optimal number of consumers and sets a quality (and price) too low compared to the social optimum. Consequently, the information (that forms platform quality) consumers release is too low. Although quantity is optimal, the welfare loss is driven by a distortion of quality (Spence distortion).

[16] As an extension of the model, it would be interesting to have two groups, informed and uninformed.

Figure 2.3: Welfare and externality relationships in different models.



The $\{low, high\}$ pairs used to evaluate the welfare functions are: $\beta = \{\frac{1}{10}, \frac{9}{10}\}$ and $\bar{\theta} = \{\frac{6}{5}, \frac{19}{10}\}$

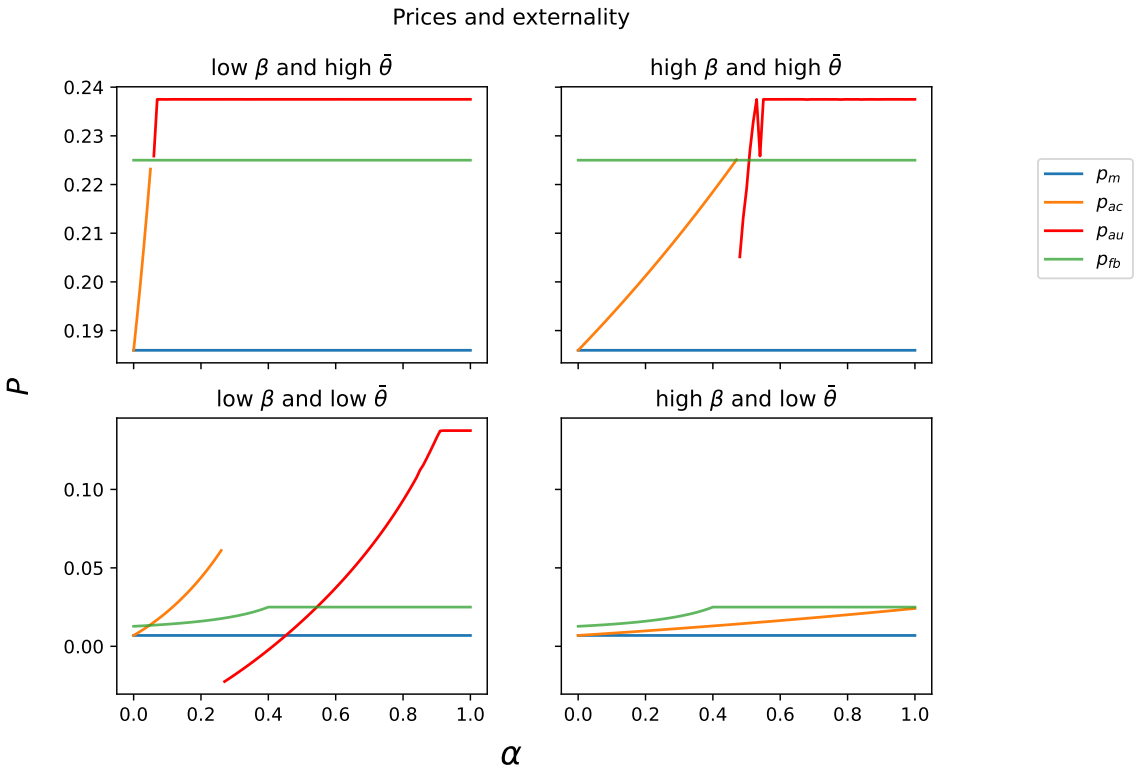
However, if consumers are aware of the externality, it impacts the indifferent consumer, and after a certain level, the firm is forced to use prices as a primary revenue instrument. For sufficiently high levels of externality, the disclosure rate drops to zero, and the price set surpass the welfare-maximizing one. In this case, the monopolist provides the optimal quality but operates with the traditional price markup.

Surprisingly, despite a welfare loss from monopoly is present in both models through all parameters configurations, in the aware case a higher value of the externality is beneficial and contributes to reducing the loss, particularly when the market is richer. This is because the externality shortens the gap between the average and the marginal willingness to pay for quality and so the Spence distortion is reduced and eventually approaches zero. In these kind of markets, the monopolist that faces aware consumers quickly switches to an uncovered market configuration that uses the price channel and so the welfare loss that arises is driven by market power exploited in the form of higher prices.

We collect the above results in the following Proposition:

Proposition 4. *Without the externality, the monopolist always under-supplies privacy and extracts consumer welfare through too high disclosure rates. In the unaware case, this distortion is aggravated by the negative externality. On the other hand, when*

Figure 2.4: Price and α



The $\{low, high\}$ pairs used to evaluate the welfare functions are: $\beta = \{\frac{1}{10}, \frac{9}{10}\}$ and $\bar{\theta} = \{\frac{6}{5}, \frac{19}{10}\}$

consumers are aware of the negative externality caused by user data correlation, the monopolist is forced to switch to a price-funded business model, and the welfare loss is reduced.

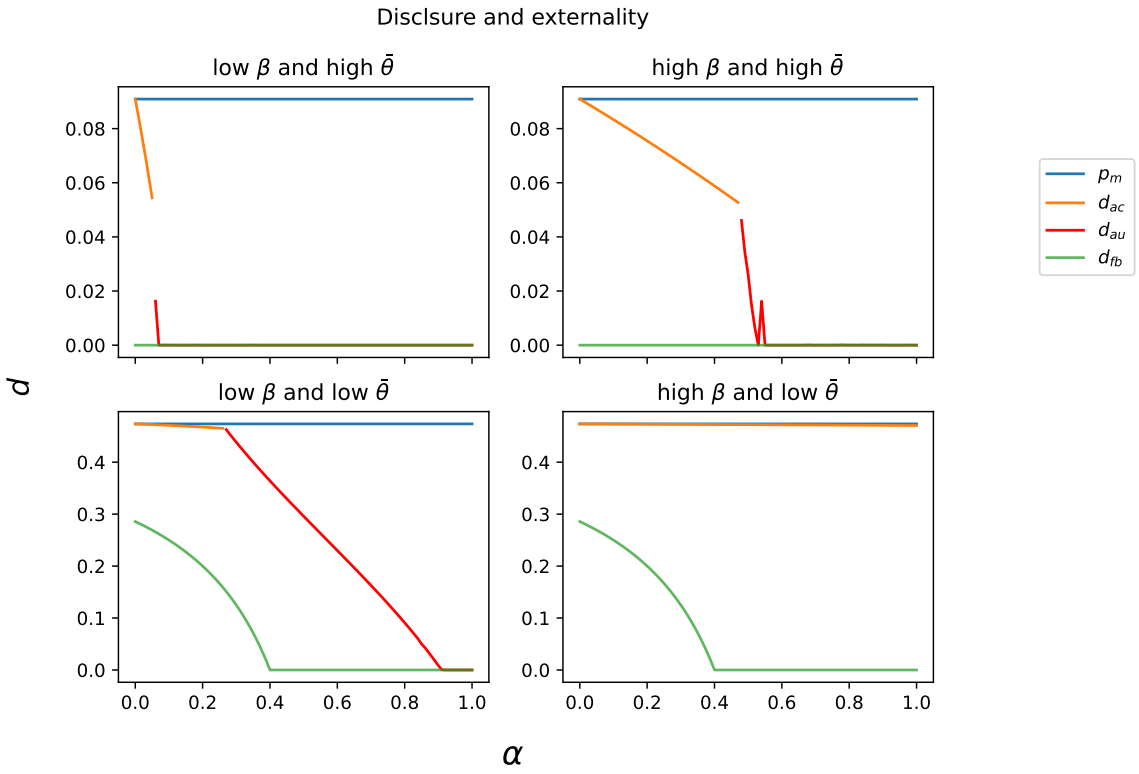
The proof of this result follows directly from the comparison of the welfare functions and the results of Proposition 3, Remarks 1 and 2, and Lemmas 1 and 2.

This implies that as long as an outside option is present ($\beta \neq 1$), according to this model, an effective policy to deal with informational externalities of this type would not be to de-correlate data (as proposed in Acemoglu et al. (2022)), but to raise consumers awareness and let the firm decide its optimal rate. Although this strategy would not represent the first best, it would reduce the welfare loss at a lower cost for regulatory agencies and consumers in richer markets.

2.7 Conclusions

Most consumers use digital services characterized by a zero monetary price and a transfer of users' personal data to the firms. In fact, 7 out of the 10 largest global companies in 2018 provided zero prices products (Mancini and Volpin, 2018). Additionally, data is correlated among consumers, and since this correlation allows firms to estimate

Figure 2.5: Disclosure rate and α



The $\{low, high\}$ pairs used to evaluate the welfare functions are: $\beta = \{\frac{1}{10}, \frac{9}{10}\}$ and $\bar{\theta} = \{\frac{6}{5}, \frac{19}{10}\}$

users' non-shared personal information, a negative externality arises (Acemoglu et al., 2022; Choi, Jeon, and Kim, 2019). In this work, we combine this externality with the facts that data disclosure and information exchanged through a platform represent an element of quality for which consumers have different willingness to pay.

We introduce the informational externality in the monopolist model of Casadesus-Masanell and Hervas-Drane (2015) that studied the firm's pricing and data disclosure decisions when personal information is an element of platform vertical differentiation. We build two variants of the model: in one, consumers are unaware of the externality at the joining stage, while in the other, they perfectly anticipate it at the joining stage, and so we include the externality in a similar fashion to the Lambertini and Orsini (2001) network effect.

We then use the models to study the welfare loss arising from monopoly, the impact of the externality on welfare, and the effect of consumer awareness. In order to do so, we had to shut out the effect of prices on the platform's advertising side, which we model as a perfectly competitive market. Additionally, we had to limit the analysis to cases where the support of the distribution is fixed at one, and there is only a monopolistic firm.

With these limitations in mind, our model delivers intriguing results that can be compared to the similar model of Choi, Jeon, and Kim (2019). They analyze the

case of data collection by a monopolist where consumers are heterogeneous for their base valuation of the service but are homogeneous for the impact of privacy features on utility. One of the results of their paper is that the welfare loss is driven by the difference in social marginal cost and private marginal cost, which is, in turn, mainly determined by the nuisance of data collection on non-users. As a result, the monopolist ends up over-collecting data and serving too many consumers.

Instead, in our model, consumers' preferences are heterogeneous for both the service value and the privacy cost, which are, however, perfectly correlated: a consumer with a higher service valuation is also willing to pay more to preserve his privacy. Additionally, consumers endogenously set the level of information released to the firm that collects all the data and decides the fraction that will disclose.

Differently from Choi, Jeon, and Kim (2019), in our model, the welfare loss arises even when non-users are not impacted and when the optimal number of consumers is served: the monopolist exerts its market power by under-providing privacy, that is, by distorting downward the quality level (Spence, 1975). Regardless of the externality, when platforms make revenues by trading personal data, they exploit market power not through prices but through disclosure (or collection) rates, even if the price consumers pay is close to zero or negative, and they have no market power in the ads markets. These results support the practice of using a quality reduction test (SSNDQ) for defining markets instead of the traditional price increase test (SSNIP) when the firms make revenues from data disclosure as proposed for zero-priced markets in Mancini and Volpin (2018).

The comparison of the first best and the model in Section 2.2 and Section 2.5 shows that when we introduce the externalities and consumers have no outside option or are unaware of them at the joining stage, then the Spence distortion is further aggravated, and the only two possibilities to restore welfare may be a Pigouvian tax on data sales or a minimum quality standard with a cap on maximum disclosure rates.

From the same comparison, we also show that despite the negative effect of the externality on consumers' utility, if consumers consider this at the joining decision, its presence increases welfare in the market compared to the no externality case. Therefore, de-correlating data is sub-optimal to spreading awareness about the externality in some cases. When the externality is strong enough, and consumers have an outside option, the sensitivity to disclosure becomes higher than the sensitivity to prices, and a marginal increase in the disclosure generates a higher loss in demand than a marginal increase in prices. Therefore, without strategic effects on the advertising side, the monopolist switches to the price channel eliminating the Spence distortion by setting a zero disclosure policy. This result has implications for regulation as making consumers aware may represent a cost-efficient strategy to increase welfare.

References

Acemoglu, D. et al. (2022). "Too much data: Prices and inefficiencies in data markets". In: *American Economic Journal: Microeconomics* 14.4, pp. 218–256.

Acquisti, A., L. Brandimarte, and G. Loewenstein (2015). "Privacy and human behavior in the age of information". In: *Science* 347.6221, pp. 509–514. ISSN: 10959203. DOI: 10.1126/science.aaa1465.

Bian, B., X. Ma, and H. Tang (2021). "The supply and demand for data privacy: Evidence from mobile apps". In: *Working paper available at SSRN*.

Bloch, F. and G. Demange (2018). "Taxation and privacy protection on Internet platforms". In: *Journal of Public Economic Theory* 20.1, pp. 52–66. DOI: https://doi.org/10.1111/jpet.12243. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jpet.12243. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jpet.12243.

Bourreau, M., B. Caillaud, and R. De Nijs (2018). "Taxation of a digital monopoly platform". In: *Journal of Public Economic Theory* 20.1, pp. 40–51.

Casadesus-Masanell, R. and A. Hervas-Drane (2015). "Competing with privacy". In: *Management Science* 61.1, pp. 229–246.

Choi, J. P., D.-S. Jeon, and B.-C. Kim (2019). "Privacy and personal data collection with information externalities". In: *Journal of Public Economics* 173, pp. 113–124.

Kumar, R. et al. (2022). "A Large-scale Investigation into Geodifferences in Mobile Apps". In: *Proceedings of the 31st USENIX Security Symposium*.

Kummer, M. and P. Schulte (2019). "When private information settles the bill: Money and privacy in Google's market for smartphone applications". In: *Management Science* 65.8, pp. 3470–3494.

Lambertini, L. and R. Orsini (2001). "Network externalities and the overprovision of quality by a monopolist". In: *Southern Economic Journal* 67.4, pp. 969–982.

Lefouili, Y. and Y. L. Toh (2017). "Privacy regulation and quality investment". In:

Mancini, J. and C. Volpin (2018). "Quality Considerations in Digital Zero-Price Markets: OECD Background Paper". In: *DAF/COMP (2018)* 14.

Spence, A. M. (1975). "Monopoly, quality, and regulation". In: *The Bell Journal of Economics*, pp. 417–429.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Appendix A

Appendix A

A.1 Endogeneity and limitations

A.1.1 Reverse causality

On the potential reverse causality issue, De Cornière and Taylor (2023) working paper provides the necessary conditions for an influence of data on market structure.

They propose a general model of competition in utility to show that the necessary condition for having more concentration is that data is unilaterally pro-competitive (UPC), and this happens when more data (or more informative data) shifts the firms' reaction function on the right.

In their model, the reaction function is shifted by more or better data (δ_i) when:

$$\frac{\partial^2 \pi_i}{\partial u_i \partial \delta_i} = \frac{\partial D_i(u_i, \mathbf{u_{-i}})}{\partial u_i} \frac{\partial r(u_i, \delta_i)}{\partial \delta_i} + \frac{\partial^2 r(u_i, \delta_i)}{\partial u_i \partial \delta_i} D_i(u_i, \mathbf{u_{-i}}) > 0, \tag{A.1}$$

where the function $D(.)$ represents the demand, and the function $r(.)$ represents the revenue function.

If equation A.1 is met, additional data will prompt the company to offer a greater level of utility through the first term known as the "markup effect." This effect is always positive and is not offset by the second term, which is called the "surplus extraction effect." The "markup effect" refers to the additional profit earned from an extra consumer. In contrast, the "surplus extraction effect" represents the opportunity cost of providing utility to consumers.

For instance, if the firm can extract surplus by showing more (or better targeted) ads, then the term $\frac{\partial^2 r(u_i, \delta_i)}{\partial u_i \partial \delta_i} D_i(u_i, \mathbf{u_{-i}})$ may compensate the incentive of the firm to offer higher utility and poach consumers, increasing market shares. If this statement were accurate, the firm would have a greater incentive not to increase the utility offered, causing data uses that extract more surplus to suffer less from reverse causality.

The proposed applications of this model distinguish different data uses by the

magnitude of the second term: for example, data for product personalization would impact next-period market concentration because the surplus extraction term is zero (data is UPC). However, other data uses, such as price discrimination, incentivize the firm to offer lower utility (due to the surplus extraction term), so data would not be UPC.

Since no exogenous variations are available, I exploit the richness of the Apple App Store information, which, differently from the Play Store, enables distinguishing among different data uses. Consequently, I form some hypotheses based on the general model of De Cornière and Taylor (2023) that provides the necessary conditions for data influence on market structure.

Their model shows that a necessary condition for having higher concentration is that more (or better) data shifts the firms' reaction function on the right: say that a social network uses data for feed personalization at t_0, then by increasing perceived quality (or addictiveness) of the platform it will start with a competitive advantage at t_1.

With this use, the unilateral effect of data is the one that dominates in the long run. Therefore, the firm could sell the same quantity at a higher price (data is unilaterally pro-competitive or UPC). On the other hand, under the assumption that more data reduces utility (such as in some cases price discrimination or when privacy preferences are strong), the firm would not build this competitive advantage, and data cannot cause the increase in market concentration (data is unilaterally anti-competitive or UAC).

Therefore, if data used for product personalization, app functionality, and app analytics are data uses unilaterally pro-competitive (UPC), then they cause an increase in concentration, and there would be a positive effect of these data uses on concentration measures. [1]

On the other hand, data used for tracking the user may activate intrinsic privacy concerns or instrumental valuations to privacy that may make this category (UAC) (Lin, 2022; Tsai et al., 2011). Thus this indicator would not directly cause an increase in market concentration, and the impact of the reverse causal link may be attenuated.

However, we need to discuss a further issue that may arise: simultaneity in the choice of the sections of the privacy summary.

A.1.2 Simultaneity bias

If, as proposed in De Cornière and Taylor (2023) and discussed in the previous section *Data Used to Track You* is not UPC, then the direct effect of this indicator

[1] With the dual instrument of prices and ads third-party advertising and developer advertising also fall into the UPC category De Cornière and Taylor (2023)

on market shares is alleviated, and the estimation would not suffer from direct reverse causality.

A mix of simultaneity and omitted variables may bias the estimates presented, and the model is valid only if the choice of the fields in "*Data Used to Track You*" is independent of the one for the fields of "*Data Linked to You*" (and "data not linked to you").

Suppose the two were not independent and were instead taken according to the models:

$$u2tu = \alpha_1 + \gamma_1 l2u + \beta_1 ms + u_1, \tag{A.2}$$

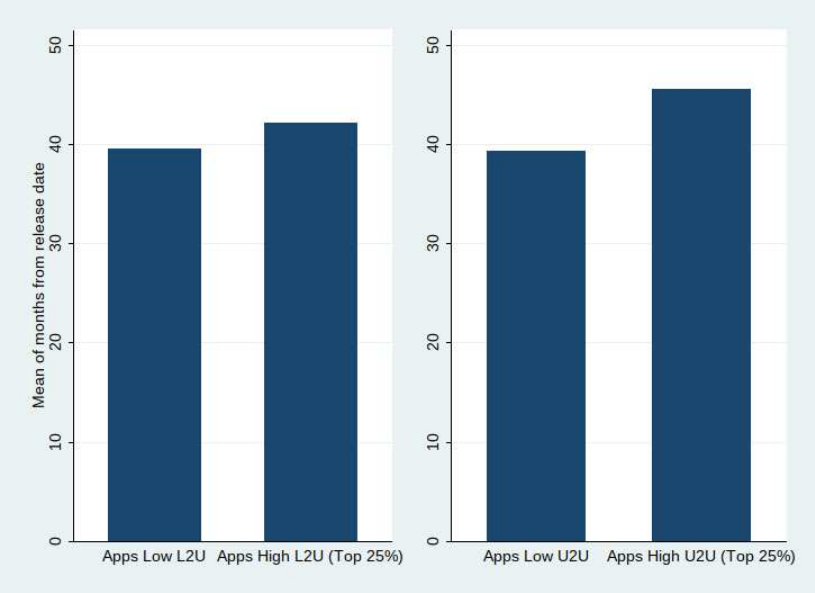
$$l2u = \alpha_2 + \gamma_2 u2tu + \beta_2 ms + u_2, \tag{A.3}$$

the model represented in eq (1.2) would suffer from omitted variable bias due to the non-inclusion of the $l2u$ indicator. Additionally, including $l2u$ would not be a solution because through the parameter γ_1, the error term of the resulting model would become a linear combination of the γs parameters and the estimates would be biased. Finally, even if γ_2 is equal to zero, we need to have $corr(u_1, u_2) = 0$ to have unbiased estimates, and the only way to solve this problem would be a structural model that describes data uses, a truly exogenous instrument or a within estimator (Wooldridge, 2015).

It is believed that the use of *L2U* and *U2TU* data involve different strategies that are independent of each other. Two reasons support this. Firstly, these indicators have a weak positive correlation, as the Pearson correlation coefficient falls between 0.2-0.3. Secondly, theory indicates that these data uses may be employed at different stages of an app's life cycle. During the development stage, data is used to build features and improve the offering through product personalization, app functionality, and analytics. However, when the app reaches maturity and gains market share, the larger installed base guarantees a higher marginal value for each data unit. At this stage, the app can infer more information about consumers, and data trades become more profitable. Therefore, it is expected that "*Data Linked to You*" and "*Data Not Linked to You*" will be reported in the early stages of the app, while "*Data Used to Track You*" will only be added when the app grows in popularity and starts to extract surplus. This is also supported by a dummy variable indicating whether the app is an entrant (less than one-year-old), which has a negative and significant estimated coefficient in the case of the *U2TU* regression, while a positive significant estimated coefficient in the case of *L2U*.

Furthermore, descriptive pieces of evidence reported in Figure A.1 weakly indicate that this may be the case: by looking at the average age of apps in the last quartile of *Data Used to Track You* and last quartile of data linked to you we can see that those that have more elements in *U2TU* are on average slightly older.

Figure A.1: Data used and app's maturity



Additionally to this figure and the coefficient of app maturity in the regression, the sequentiality of the decision is also reflected by the sample summary statistics. By splitting entrants from the established firms, it emerges that while the established firms use only 18% more data items linked to the user than an entrant, they collect about 50% more items used to track consumers than an entrant. Additionally, by further decomposing the differential of the "*Data Linked to You*" indicator by its purposes and investigating those that change the most among apps of different maturity, we observe the most significant change in that third-party advertising and developer advertising. This result is coherent with a gradual shift in business model as the app gains a more significant installed base. Following entrants for an extended period would allow future studies to confirm this hypothesis and provide unambiguous evidence of a switch in the business model or proof of the survival thesis.

A.1.3 Economies of scale and impact of privacy preferences

A further concern that would motivate the simultaneity of the system (A.3) is the presence of economies of scale in using the same data field. Once an app collects the browsing history for product personalization, the extra technical cost of collection and security of that item associated with selling it (*Data Used to Track You*: shared to third parties or data brokers) is basically null.

However, the decrease in demand due to privacy preferences that the app may face if they include the item in the '$U2TU$' may represent an impactful opportunity cost. In this case, the high opportunity cost may balance the cost savings due to economies of scope and may be highly relevant in deciding which items to include.

If none of these assumptions hold, I propose a robustness test by including the competitors' average level of "*Data Linked to You*" in the regression to proxy the potential simultaneity link between the two variables. This instrument would not be endogenous in the case that the (infinitesimal) firm is a "privacy taker" in the clusters, an assumption that may hold only in some clusters. The coefficient for this proxy is positive and significant, but the change in the main results for the market share effect is low, and the main results hold.

A.1.4 Impact of data on updates

Even if the indicators are chosen independently, data may impact developers' innovation ability, which in turn may impact market shares and market power. Consequently, in the estimates presented until now, there could be an upward bias due to the potential positive correlation of updates with downloads and market shares (Comino, Manenti, and Mariuzzo, 2019).

Apps that obtain more data can update more often or raise the quality of updates, and this would result in an increase in the quantity downloaded and in the rating share of the app. To test for this bias, we added the additional covariates of update frequency and count of updates and found that the results do not change consistently.

Additionally, for robustness, I provide estimated auxiliary regression models using as dependent variables updates indicators and as main explanatory variables the type of data uses. The view provided by these auxiliary regressions is coherent with the fact that "*Data Used to Track You*" is not used in the updating process. Although "*Data Linked to You*" and "*Data Not Linked to You*" indicators positively affect the probability of seeing a version change, the total number of updates and increase the update frequency (updates per month), "*Data Used to Track You*" indicator is either not significant or impacts in the opposite direction updates. To my knowledge, this is the first attempt to measure the impact of various data applications on online market shifts. Albeit in an early and still embryonic stage results are presented in Appendix A.4.

A.2 Sensitivity to Market Definition

A.2.1 Sensitivity to Resolution Parameter with YMAL network

Table A.1 shows how the resolution parameter impacts the results. Modularity is maximized when the resolution parameter is equal to one. However, one can adjust this parameter to identify smaller, more homogeneous communities. Despite the importance of this parameter, this is not reported in Kesler, Kummer, and Schulte (2019).

I show that by increasing the resolution on this particular network, the magnitude of the effects halves when reaching R=20 and eventually becomes non-significant when reaching R=70.

In the case of the network used in this article, the value of R=10 already gave good communities. However, conservatively, the results have been reported for R=20 because after this threshold, the communities' quality started degrading, and large hubs started forming distinct clusters. As an example, above such resolution, I observed that competing apps, such as "Pandora: Music & Podcasts" and "Spotify - Music and Podcasts", fell into different clusters with no major competitor.

A.2.2 Alternative market definition

General market definition considers three main elements: demand substitution, supply substitution, and entry/expansion patterns Motta (2004). The first of these elements is the one that the YMAL section is most suited to capture because it is drawn directly from the purchasing patterns of consumers. Description analysis, however, integrates multiple dimensions by considering the app's similarities in the feature space from both a demand-side and a supply-side perspective. Apps with low substitutability on the demand side may have high substitutability on the supply side, and the feature set may capture this pattern. As an example, on the demand substitutability side, an app that reminds watering the garden, such as 'WaterMe - Gardening Reminders', is very different from an app, such as 'Water Reminder - Daily Tracker', that tracks the amount of water you drink and reminds drinking water. However, on the supply side, the code may be sufficiently similar to quickly adapt a version of one app to the other market, and therefore, considering the descriptions may allow capturing 'small entry' by developers.[2]

Thus, if we consider an app as a bundle of features and assume that the descriptions tend to express them, we can exploit them to define markets by obtaining a *term-document-matrix* (TDM), apply a cosine similarity measure to obtain a document-document similarity matrix and model it as a network on which we can perform the modularity maximization through the Louvain algorithm.

Methodology

Here, I integrate the approach of Hoberg and Phillips (2010), Hoberg, Phillips, and Prabhala (2014), and Pellegrino (2023) with applied network analysis based on modularity maximization also used in Kesler, Kummer, and Schulte (2019).

After pre-processing the descriptions by removing all URLs, punctuation, and sec-

[2]Focusing on the categories (Lifestyle in the first case and Health & Fitness) instead, would not capture either of these two elements because of the app heterogeneity you find in each category

Table A.1: Sensitivity of the Seller FE model to the resolution parameter and market definition

	R 3	R 10	R 15	R 20
	(1)	(2)	(3)	(4)
HHI	-0.008	0.025**	0.009	0.005
	(0.010)	(0.009)	(0.007)	(0.007)
Log of market share	1.039**	0.812***	0.574***	0.443***
	(0.320)	(0.167)	(0.133)	(0.104)
Log of rating count	0.019***	0.019***	0.019***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)
Price dummy, =1 if price>0	-0.210***	-0.210***	-0.210***	-0.210***
	(0.012)	(0.012)	(0.012)	(0.012)
Dummy variable for in-app purchases	0.047***	0.047***	0.047***	0.047***
	(0.006)	(0.006)	(0.006)	(0.006)
Numeric count of the languages of the app	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Dummy variable, =1 if mac version exist	-0.012**	-0.012**	-0.012**	-0.012**
	(0.004)	(0.004)	(0.004)	(0.004)
Log(N. apps) by seller and wave	-0.020**	-0.020**	-0.020**	-0.020**
	(0.007)	(0.007)	(0.007)	(0.007)
Age rating (PEGI): Baseline 4+				
Factor variable, 4+ 9+ 12+ or 17+=9	0.016	0.016	0.016	0.016
	(0.009)	(0.009)	(0.009)	(0.009)
Factor variable, 4+ 9+ 12+ or 17+=12	0.011	0.012	0.012	0.012
	(0.006)	(0.006)	(0.006)	(0.006)
Factor variable, 4+ 9+ 12+ or 17+=17	0.015***	0.015***	0.016***	0.016***
	(0.004)	(0.004)	(0.004)	(0.004)
App Maturity (months old): Baseline Very young (0-12)				
Young (13-21m/o)	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Mature (22-37m/o)	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Very Mature (38-66m/o)	0.000	0.000	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
Veteran (67-121m/o)	0.009*	0.009*	0.009*	0.009*
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.166***	0.164***	0.165***	0.166***
	(0.010)	(0.010)	(0.010)	(0.010)
Observations	2317525	2317525	2317525	2317525
R^2	0.847	0.847	0.847	0.847

Note: this regression includes category and seller dummies in the pooled OLS regression. The number of observations is lower than the full dataset because singletons are automatically dropped by the Stata command 'reghdfe,' which would otherwise artificially reduce standard errors.
Standard errors in parentheses. Significance levels are: *** p<0.001, ** p<0.01, * p<0.05

tions not in English characters, all descriptions of the non-game apps have been tokenized to form a Term-Document-Matrix (with dimensions 50063-754430 and a maximal term length of 60).

To build this matrix, Hoberg and Phillips (2010) utilizes a binary weighting function with a ceiling to limit the analysis to rare words (words appearing in more than 5% of the document were excluded).

Using a binary weighting function strongly reduces the accuracy of the analysis because it does not account for the frequency of that word in the document or the frequency in the whole document set. Additionally, it has to rely on a subjective threshold of 5%.

Instead, using *Term Frequency inverse Document Frequency* (TFiDF) weights and assigning the weight of each cell in the matrix according to the formula

$$v_{ij} = TF_{ij}/DF_i,$$

where TF is the frequency of term i in document j and DF is the frequency of term i across all documents, allows to give more weight to rare words (in the documents set) that repeatedly appear in a single document. [3]

After obtaining the TDM, as a last step to obtain a similarity matrix, I use a version of cosine similarity comparable to Hoberg and Phillips (2010):

$$C_{n,n} = M^T M,$$

where M is the Term-Document Matrix.

The resulting $C_{n,n}$ is a document-document matrix of similarity indexes that go from 0.3 to one. All the values of similarity below 0.3 have been deleted to ease computations and not to employ dense matrices that would overload the available memory. This 0.3 threshold was the lowest I could reach by hardware limitation. Unfortunately, with an extensive similarity matrix like the one used, the resulting network from the dense matrix multiplication would be larger than the 200Gb of ram available in the hardware used.

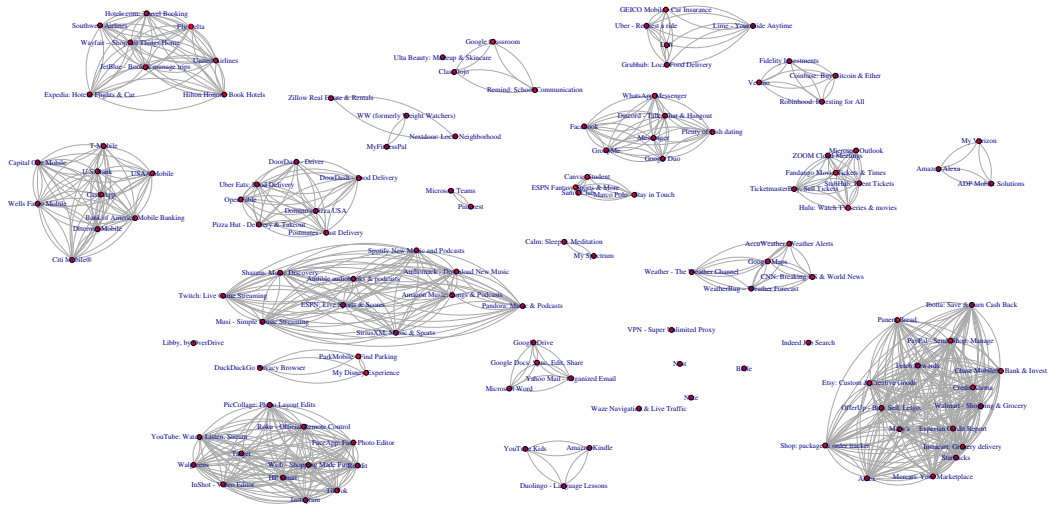
As an example, and to show this procedure's capabilities and pitfalls, Figure A.2 shows the unweighted associated network for the top Apps of the Apple Store.

I used a static version of the market definition to simplify the process and prevent description changes from affecting the data. This involved combining all waves and merging app descriptions by app while removing duplicate terms. Although this method may group apps with added or removed functions, it is still preferred to a dynamic market definition due to the short observation period. Finally, I applied

[3] Additionally, every column vector has been standardized to have a unit length.

modularity maximization using the Louvain algorithm to identify communities within the network. For more information on the procedure and mathematical aspects, please refer to in-text Section 1.2.3.

Figure A.2: Network based on description similarity



Note: The network has been filtered with 0.3 as a minimum score to have a link. Figure obtained by considering top non-gaming apps to test the correct classification of different apps. Despite a good overall accuracy, the algorithm shows that productivity apps tend to be thrown together despite being in different sectors.

Results

The results presented in Table A.2 are consistent with the results of the YMAL network. The magnitude is slightly lower than the estimates in the text, but the sign and significance are unvaried.

Notice that the number of observations is lower because apps with too short descriptions or zero characters descriptions were excluded. Their inclusion would imply many single-app clusters with unitary market share but are in markets of less than 1000 ratings. These 1035 apps (7000 observations) were their communities' only elements. Due to these markets' competitiveness and the superstars' economics, apps with a unitary market share should be rare. Therefore, including these apps brings measurement errors that have the potential to bias the results. By including these apps, the sign of the app id fixed effects model was negative and statistically significant, and all other results were mostly unaffected.

Given these apps' negative impact on the market share coefficient, we expect that the simple exclusion introduces an attenuation bias of the positive coefficient of market

Table A.2: Sensitivity to market definition based on text analysis

	POLS	Dev. FE		Id FE	
	(1)	(2)	(3)	(4)	(5)
HHI	-0.158***	-0.023***	-0.021***	0.024**	0.024**
	(0.003)	(0.004)	(0.004)	(0.007)	(0.007)
Log of market share in clusters	0.413***	0.336***		0.119	
	(0.055)	(0.054)		(0.117)	
Categorized Share (baseline x<5%)					
0.05 < x <= 0.2			0.068**		0.037*
			(0.025)		(0.017)
0.2 < x <= 0.4			0.120**		0.030
			(0.044)		(0.017)
0.4 < x <= 0.8			0.110**		0.038
			(0.040)		(0.040)
0.8 < x <= 1			0.117**		0.066
			(0.038)		(0.066)
Log of rating count	0.033***	0.019***	0.019***	0.015***	0.015***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Price dummy, =1 if price>0	-0.142***	-0.210***	-0.210***	-0.026***	-0.026***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Dummy variable for in-app purchases	0.179***	0.047***	0.047***	0.043***	0.043***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Numeric count of the languages of the app	0.002***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dummy variable, =1 if mac version exist	-0.000	-0.012***	-0.012***	-0.016***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Log(N. apps) by seller and wave	0.020***	-0.020***	-0.020***	-0.013***	-0.013***
	(0.000)	(0.002)	(0.002)	(0.001)	(0.001)
App Maturity (months old): Baseline Very young (0-12)					
Young (13-21m/o)	0.002**	-0.000	-0.000	0.005***	0.005***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Mature (22-37m/o)	-0.016***	-0.001	-0.001	0.017***	0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Very Mature (38-66m/o)	-0.023***	0.000	0.000	0.030***	0.030***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Veteran (67-121m/o)	0.005***	0.009***	0.009***	0.047***	0.047***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age rating (PEGI): Baseline 4+					
Factor variable, 9+	0.135***	0.016***	0.016***	-0.023	-0.023
	(0.002)	(0.004)	(0.004)	(0.013)	(0.013)
Factor variable, 12+	0.063***	0.012***	0.012***	0.024***	0.024***
	(0.001)	(0.002)	(0.002)	(0.006)	(0.006)
Factor variable, 17+	0.074***	0.016***	0.015***	-0.007*	-0.007*
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Constant	0.065***	0.168***	0.168***	0.138***	0.138***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	2310689	2310689	2310689	2310689	2310689
R^2	0.122	0.847	0.847	0.955	0.955

share.[4] Results are robust to the resolution parameter.

[4]Ideally, we could re-classify them based on a scoring function that uses category and name to the most similar cluster. However, most of the insights of the analysis are the same.

A.3 Inspection of apps that changed U2TU section

The panel consists of around 2.3 million observations spanning over 6 periods and covering 434955 apps. Only 14891 changes occurred in the variable $U2TU$ over time, with 76.5% resulting in an increase in value and 23.4% in a decrease.

Table A.3: Summary Statistics Data Uses

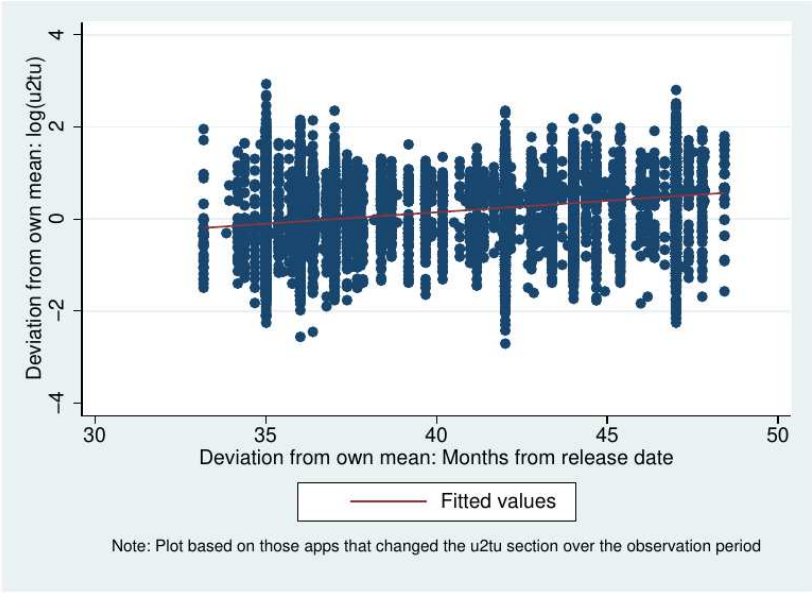
	Decrease in U2TU			Constant U2TU			Increase in U2TU		
	count	mean	sd	count	mean	sd	count	mean	sd
Count of number of packages	3483	1.077	2.510	2396025	0.542	1.815	11408	1.290	2.723
Avg. Price of the inapp purchases	3483	5.754	20.889	2396025	2.908	19.289	11408	5.273	17.819
Months from release date	3476	43.012	33.884	2386155	40.142	33.590	11402	48.146	34.632
Log of market share	3483	0.002	0.016	2396025	0.001	0.011	11408	0.002	0.018
HHI	3483	0.101	0.127	2396025	0.106	0.123	11408	0.099	0.121
Rating Count	3483	2517.478	33771.975	2396025	2073.843	95457.429	11408	3622.916	69167.021
Average stars, NA if no ratings	2435	4.266	0.939	1312602	4.217	1.053	8340	4.221	0.987
Observations	2410916								

Table A.3 displays significant differences in sample averages of crucial variables among the apps that altered the value of $U2TU$ and those that did not. After analyzing the data, it is apparent that apps with an increase in the number of items in the privacy summary are usually free, have a higher market share and rating count, are more up-to-date, older, and have more expensive in-app purchases compared to the sample average. In-app purchases seem to be correlated with the $U2TU$ variable, as the apps that experienced an increase have almost three times the sample average. The higher market share and rating count of these apps may be because they have a lower price, making them more accessible to consumers who try the app and leave reviews. It could also be due to the strategic effect of constantly updating the app. However, the fact that these apps are older raises questions about whether they have changed their business model or driven out competitors by using data. It is worth noting that there is a positive correlation between the app's maturity and the within variation of the variable $U2TU$ for apps that changed their privacy summaries (depicted in Figure A.3). This suggests that with a more extended observation period, we may gain more insights into this phenomenon. Further analysis of entry and factors impacting the success of entrants could indicate whether using data to track is a competitive tool necessary to improve the product or just a means to extract surplus.

A.4 Apps updates behavior

Apps' product innovation can be incremental, through more app updates or entirely new apps. While the latter innovation methodology is the most effective strategy for

Figure A.3: Deviation from apps mean *log(U2TU)* vs *months old*



the success of gaming apps, non-gaming applications are likelier to become killer apps if the developer constantly improves the app Yin, Davis, and Muzyrya (2014). Since I exclude gaming apps from the dataset, I focus on the update side of innovation in this section. [5]

If data is an input in the updating process, not to have endogeneity in the model for $U2TU$, we shall have that $L2U$ and $NL2U$ indicators impact updates. However, our main dependent variable should not impact them.

I use three approaches to understand the relationship between data and updates. First, I build a dummy variable to identify version changes between waves and estimate a logit model to predict the probability of observing an update. I use three data collection indexes, category dummies, app-level and developer-level controls as main regressors. Second, I estimate a Tobit model using the censored count of updates as the dependent variable. Third, I normalize the count of updates to obtain an update frequency indicator and then regress it on the data collection indexes. These three approaches complement each other and provide a comprehensive understanding of updates.

We estimate all the models with lagged values to consider the decision's sequentiality, and that data is an input for the update process. Additionally, we want to avoid the endogeneity of the regressors that may arise from the positive impact that updates have on market share by creating a buzz around the app. Consequently, the developer may also decide to increase data extraction.

[5]Furthermore, due to a lack of data on costs, I neglect another innovation aspect that may be data-driven, cost-reducing innovations. Unfortunately, since we do not observe the cost side of the market, I abstain from commenting on cost-saving innovation due to data collection.

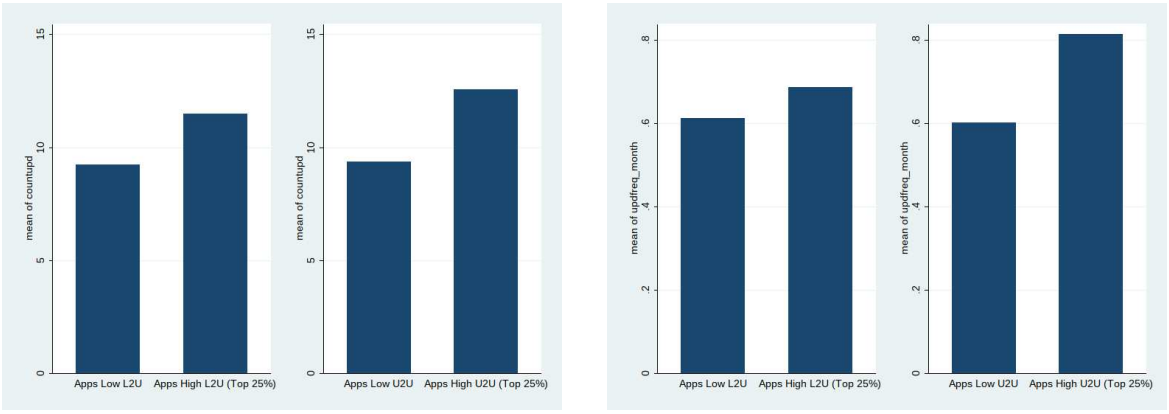
We expect that $l2u$ and $nl2u$ indicators would affect the updating probability, while $U2TU$ would not. The results of this robustness test are in column (1) of Table A.4.

Another way to look at updates would be to regress the count of updates on the data indicators. However, given that the count of updates is capped at 25, and older apps are more likely to have reached the threshold, this regression would capture other effects due to the app's maturity. Therefore, I estimate a tobit regression model on the count of updates to consider that observations are censored. We expect that the two indicators $l2u$ and $nl2u$ positively affect the app's number of updates, while $U2TU$ shall not. The results are presented in columns 5 to 7 of Table A.4. We check for robustness of this regression by introducing lagged values for the data uses indicators and two indicators of app quality: the fraction of 5-star and 1-star ratings. Given the results in Comino, Manenti, and Mariuzzo (2019), I expect the decision to update to correlate positively with the fraction of 1-star ratings and negatively with the fraction of 5-star ratings. Therefore we control for these values.

Finally, to avoid the downsides of the censored count of updates, I created a frequency of updates indicator by normalizing the count of updates by the range in which these are released. The mean of this indicator suggests that the apps in the sample released an update every 50 days, which is a value in line with Comino, Manenti, and Mariuzzo (2019). Therefore, if more data is beneficial in the innovation process, we would expect the release of more updates and a positive effect of the data indicators $l2u$ and $nl2u$ on the frequency of updates and no effect of the *Data Used to Track* consumers. Results are reported in columns 2-3-4.

A.4.1 Descriptives

Figure A.4: Updates indicators for last quartile of L2U and U2TU vs. first three quartiles



(a) Mean number of updates

(b) Mean of update frequency (upd/months)

updates Figure A.4 shows a considerable jump in the average number and frequency of updates. However, this correlation may be driven by the older age of apps with more U2TU. In fact, more mature apps also have more extended version history. Therefore, we shall control for the app's maturity in the regression and check whether the effect is robust to this factor.

A.4.2 Results

The analysis shows that in all three regressions, $L2U$ and $NL2U$ positively impact the probability of a version change and are associated with a higher frequency of updates, resulting in more updates. On the other hand, the impact of the variable $U2TU$ is not statistically significant or has a more negligible impact on these indicators. It is noteworthy that the coefficient of $U2TU$ (lagged in columns (1) and (7) and not lagged in columns (5) and (6)) has a negative and significant sign. This suggests that *Data Used to Track* is associated with a lower chance of observing a version's change and a lower cumulative number of updates, all else being equal. One possible explanation is that apps with higher *Data Used to Track* individuals have greater market power and do not need to rely as much on updates to drive downloads. These results confirm the findings in the text.

When data is used to develop a better app, we expect a positive correlation between the shares and these indexes of data usage. All these descriptive pieces of evidence indicate that in a structural model, the equations would follow a pattern such as:

$$upd = g(\text{l2u/nl2u}, \text{ability}, \#\text{ of downloads}, \ldots), \tag{A.4}$$

$$ms = f(\text{updates}, \text{features}, \text{price}, \ldots), \tag{A.5}$$

$$u2tu = z(\text{share}, \#\text{ of packages}, \text{price}, \ldots), \tag{A.6}$$

with or without the reverse causality element for $L2U/NL2U$, as long as there is independence between the choices of the privacy summary section, and $U2TU$ is UAC, the results in the regression (1.2) could be interpreted as a signal of the effect of market power.

To check for the consistency of the result in Table 1.4, I also introduced the number of updates and whether an app has in-app purchases or not in the regression. The results show that despite both variables have a positive (and significant) effect on the amount of *Data Used to Track* consumers, the market share variable estimated coefficient does not substantially change in magnitude, sign, or significance.

Therefore, if $U2TU$ and $L2U$ are not co-determined, as in (A.3), the findings of this study may indicate the impact of market power on data markups. However, if they are co-determined, it is advisable to interpret the results with caution. To address

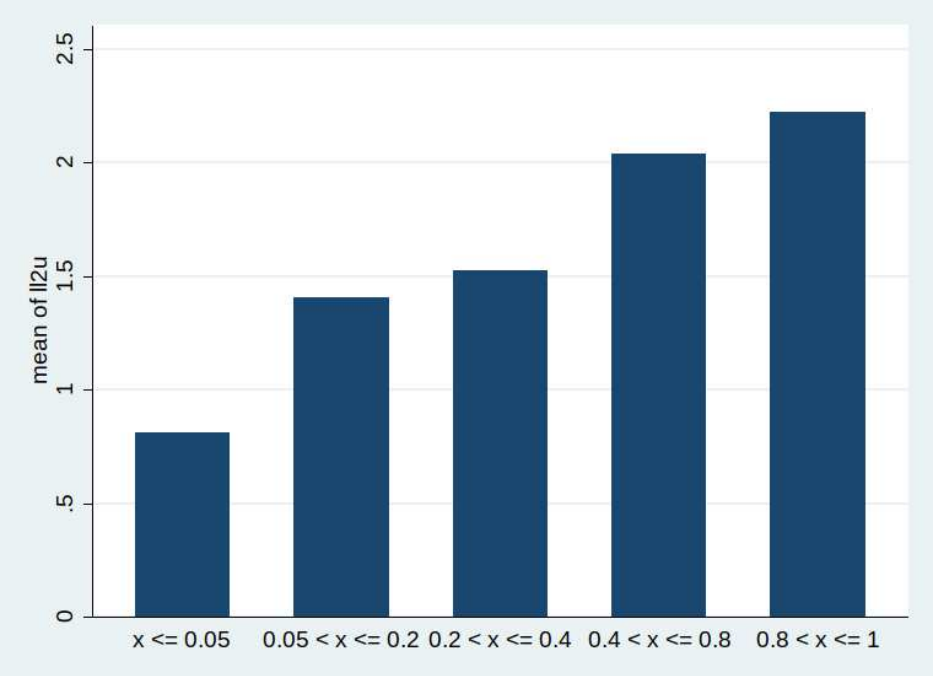
Table A.4: Selected coefficients for model on update frequency and count of updates

	Dummy = 1 if vers. change	Upd. Freq. (month)			Count update		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
main							
Lag of Log(data linked to you)	0.141***			0.042***			0.472***
	(0.003)			(0.005)			(0.046)
Log(data linked to you)		0.025***	0.042***		0.457***	0.450***	
		(0.003)	(0.005)		(0.043)	(0.045)	
Lag of Log(data not linked to you)	0.152***			0.028***			0.941***
	(0.004)			(0.006)			(0.033)
Log(data not linked to you)		0.012***	0.025***		0.879***	0.918***	
		(0.004)	(0.006)		(0.032)	(0.033)	
Lag of Log(data used to track you)	-0.029***			0.002			-0.362***
	(0.007)			(0.012)			(0.078)
Log(data used to track you)		0.013	0.002		-0.384***	-0.354***	
		(0.008)	(0.012)		(0.076)	(0.079)	
Dummy in-app purchases, =1 if at least one package	0.055***	0.121***	0.146***	0.146***	1.054***	1.021***	1.029***
	(0.009)	(0.010)	(0.015)	(0.015)	(0.083)	(0.089)	(0.089)
Price dummy, =1 if price>0	-0.422***	-0.010	-0.004	-0.004	-1.646***	-1.546***	-1.532***
	(0.017)	(0.010)	(0.014)	(0.014)	(0.107)	(0.108)	(0.109)
Log(N. apps) by seller and wave	-0.101***	0.045***	0.025	0.025	-0.643***	-0.565***	-0.566***
	(0.002)	(0.007)	(0.016)	(0.016)	(0.063)	(0.065)	(0.066)
Log(rating count)	0.162***	0.065***	0.071***	0.070***	1.519***	1.474***	1.470***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.017)	(0.017)	(0.017)
Log of % of 5 stars ratings	-0.055***		0.042***	0.042***		-1.096***	-1.097***
	(0.009)		(0.008)	(0.008)		(0.072)	(0.073)
% of 5 stars ratings		0.022***			-1.095***		
		(0.006)			(0.072)		
Log of % of 1 stars ratings	-0.046***		-0.019*	-0.019*		-0.283***	-0.277***
	(0.017)		(0.011)	(0.011)		(0.082)	(0.082)
% of 1 stars ratings		-0.041***			-0.272***		
		(0.007)			(0.078)		
Average count of updates by developer (excluding that app)	0.049***				0.434***	0.400***	0.399***
	(0.001)				(0.007)	(0.007)	(0.007)
App Maturity (baseline 0-13 m/o)							
Young (13-21m/o)	-0.397***	-0.175***	-0.263***	-0.262***	3.008***	3.825***	3.836***
	(0.009)	(0.003)	(0.009)	(0.009)	(0.031)	(0.040)	(0.040)
Mature (22-37m/o)	-0.539***	-0.325***	-0.407***	-0.406***	5.629***	5.927***	5.940***
	(0.010)	(0.005)	(0.009)	(0.009)	(0.047)	(0.055)	(0.056)
Very Mature (38-66m/o)	-0.620***	-0.392***	-0.462***	-0.462***	7.140***	7.333***	7.348***
	(0.011)	(0.006)	(0.010)	(0.010)	(0.118)	(0.119)	(0.120)
Veteran (67-121m/o)	-0.699***	-0.421***	-0.501***	-0.500***	10.074***	10.431***	10.449***
	(0.011)	(0.008)	(0.013)	(0.013)	(0.084)	(0.098)	(0.097)
Constant	-2.011***	0.637***	0.695***	0.693***	0.366***	0.391***	0.374***
	(0.038)	(0.011)	(0.022)	(0.022)	(0.126)	(0.126)	(0.128)
sigma							
Constant					7.180***	6.854***	6.850***
					(0.032)	(0.030)	(0.029)
Observations	715721	2401033	827924	827924	2401033	827924	827924
R^2		0.723	0.787	0.787			
Pseudo R^2	0.065				0.089	0.093	0.093

Note: The first column indicates the logit model for the probability of a version change, with category dummy. The columns (2-3-4) come from a seller-level fixed effects regression. In contrast, the coefficients in columns (5-6-7) come from a Pooled Tobit regression to account for the upper censoring of the update history. Significance levels are: *** p<0.001, ** p<0.05, * p<0.10

potential endogeneity issues, an exogenous variation or a structural model could be used to estimate the effect of market share on *Data Used to Track*.

Figure A.5: Log of *L2U* by category of market share



A.5 Regression for data linked to you indicator

Some descriptive evidence is in A.5. It shows a positive relationship even stronger than the one for *U2TU*.

This is confirmed by the regression analysis in Table A.5. However, given that this includes functionalities that improve the product the estimated coefficient are likely biased. Computing the direction of the bias is particularly complex in this setup and it would depend on the strength of each side correlation.

Table A.5: Regression analysis with the log. of data linked to you as dependent variable

	Dep. Var $log(l2u)$				
	R 3	R 10	R 15	R 20	Radius
	(1)	(2)	(3)	(4)	(5)
hhi	0.004	0.047**	0.031*	0.011	
	(0.021)	(0.020)	(0.016)	(0.014)	
Rating share in the radius of similar apps					-0.020
					(0.016)
Log of market share in clusters	0.433	0.507**	0.493**	0.418**	
	(0.357)	(0.233)	(0.220)	(0.167)	
Log of rating count	0.052***	0.051***	0.051***	0.051***	0.052***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Dummy in app purchases, =1 if at least 1 package	0.043***	0.043***	0.043***	0.043***	0.043***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Price dummy, =1 if price>0	-0.167***	-0.167***	-0.167***	-0.167***	-0.167***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Age rating (PEGI): Baseline 4+					
Factor variable, 9+	0.035**	0.035**	0.035**	0.035**	0.035**
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Factor variable, 12+	0.083***	0.083***	0.083***	0.083***	0.083***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Factor variable, 17+	0.078***	0.078***	0.078***	0.078***	0.078***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Number of updates per month	0.018***	0.018***	0.018***	0.018***	0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Numeric count of the languages of the app	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Dummy variable, =1 if mac version exist	-0.046***	-0.046***	-0.046***	-0.046***	-0.046***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
App Maturity (months old): Baseline Very young (0-12)					
Young (13-21m/o)	-0.005**	-0.005**	-0.005**	-0.005**	-0.005**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mature (22-37m/o)	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Very Mature (38-66m/o)	-0.010*	-0.009*	-0.009*	-0.009*	-0.010*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Veteran (67-121m/o)	0.010	0.010	0.010	0.010	0.010
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Log(N. apps) by seller and wave	-0.084***	-0.085***	-0.085***	-0.084***	-0.084***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	0.853***	0.850***	0.851***	0.853***	0.852***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Observations	2320133	2320133	2320133	2320133	2320133
R^2	0.891	0.891	0.891	0.891	0.891

Note: This regression uses the same functional form of the seller-category fixed effect model for *Data Used to Track* consumers. The number of observations is lower than the full dataset because 5274 singletons are automatically dropped by the Stata command 'reghdfe', because they would otherwise artificially reduce standard errors and overstate significance. The standard errors reported in parentheses are clustered at the seller level to account for possible heteroskedasticity. Significance levels are: *** p<0.001, ** p<0.05, * p<0.10

Appendix B

Appendix B

B.1 Model with aware consumers

Proof. Full Market Coverage case: with $\hat{\theta} = \bar{\theta} - 1$ demand is $x^{fmc} = 1$ and the indifferent consumer needs to be located at the left of the consumer with the lowest θ. So to obtain FMC we need the indifferent condition in eq.B.1 respected and satisfied with equality:

$$\theta\frac{(1-d)^2}{4} - P - e\frac{1-d}{2}d \geq 0,$$

$$\theta^* = \frac{4\left(P + e\frac{1-d}{2}d\right)}{(1-d)^2} \leq \bar{\theta} - 1, \tag{B.1}$$

we can then solve this equation by finding the price P_c that satisfies with equality the covered market condition in (B.1):

$$P_c = \frac{1}{4}(d-1)(d(2e + \bar{\theta} - 1) - \bar{\theta} + 1), \tag{B.2}$$

with this price the profit function becomes:

$$\pi^{fmc} = \frac{1}{4}(d-1)(d(2e + \bar{\theta} - 3) - \bar{\theta} + 1), \tag{B.3}$$

we can then maximize the profit function in d and get the following F.O.C.:

$$\frac{\partial \pi^{fmc}}{\partial d} = \frac{1}{2}(d(2e + \bar{\theta} - 3) - e - \bar{\theta} + 2) = 0, \tag{B.4}$$

$$\frac{\partial^2 \pi^{fmc}}{\partial d^2} = \frac{1}{2}(2e + \bar{\theta} - 3) = 0, \tag{B.5}$$

when the second order condition is not respected, the function is convex and the profit function is maximized at the corner $d = 0$, additionally by solving the FOC for the

optimal disclosure we obtain:

$$d^*_{fmc} \begin{cases} \dfrac{e+\bar{\theta}-2}{2\,e+\bar{\theta}-3} & \text{if} \quad \bar{\theta} \leq 2-e \ \wedge \ \bar{\theta} \neq 3-2e, \\ 0 & \text{Otherwise,} \end{cases} \tag{B.6}$$

where the condition:

$$\bar{\theta} \leq 2-e \iff \alpha \leq \frac{2-\bar{\theta}}{1-\beta},$$

derives from the constraint that d^* is bounded between zero and one. Instead the second order conditions requires:

$$\bar{\theta} \leq 3-2e \iff \alpha \leq \frac{3-\bar{\theta}}{2(1-\beta)},$$

it is obvious that the first constraint is stricter than the one imposed by the SOC.

That is, when the net impact of the externality on the demand function is large enough the monopolist optimally sets the boundary solution of the zero disclosure rate: d drops to zero before the function becomes convex.

Pulling all together we have final prices and profits of Proposition 1:

$$P^*_{fmc} = \begin{cases} \dfrac{(1-e)\left(2e^2-e(\bar{\theta}-3)+\bar{\theta}-1\right)}{4(2e+\bar{\theta}-3)^2} & \text{if} \quad \bar{\theta}+e \leq 2 \ \wedge \ \bar{\theta} \neq 3-2e, \\ \dfrac{\bar{\theta}-1}{4} & \text{Otherwise,} \end{cases} \tag{B.7}$$

$$\pi^*_{fmc} = \begin{cases} \dfrac{(1-e)^2}{4(3-\bar{\theta})-8e} & \text{if} \quad \bar{\theta}+e \leq 2 \ \wedge \ \bar{\theta} \neq 3-2e, \\ \dfrac{\bar{\theta}-1}{4} & \text{Otherwise,} \end{cases} \tag{B.8}$$

as the market is covered, and the fraction of non-users is zero, consumer welfare is defined as:

$$CS = \int_{\bar{\theta}-1}^{\bar{\theta}} U_{join} dz = \int_{\bar{\theta}-1}^{\bar{\theta}} \left(\frac{z(1-d)^2}{4} - \alpha d y^* - P \right) dz, \tag{B.9}$$

which results in:

$$CS_a = \begin{cases} \frac{1}{8} & 2e+\bar{\theta}=3 \wedge e+\bar{\theta}>2 \\ \dfrac{(e-1)(-4\alpha(e+\bar{\theta}-2)+4e(e+\bar{\theta})-7e-1)}{8(2e+\bar{\theta}-3)^2}, & \text{Otherwise,} \end{cases} \tag{B.10}$$

and summing up profit and consumer surplus gives total welfare:

$$W_a = \begin{cases} \frac{1}{8}(2\bar{\theta}-1) & 2e+\bar{\theta}=3 \wedge e+\bar{\theta}>2, \\ \frac{(e-1)(-4\alpha(e+\bar{\theta}-2)+2(e+1)\bar{\theta}+3e-7)}{8(2e+\bar{\theta}-3)^2} & \text{Otherwise,} \end{cases} \tag{B.11}$$

□

B.2 Partial market coverage with aware consumers

Here the results of the simulation and a short code snippet are presented.

From the Table, we can observe the following patterns:

Table B.1: Simulation results for monopolist equilibrium with aware consumers

π_c	π_u	d_u	α	β	e	$\bar{\theta}$	Configuration
8 0.1250	0.1111	0.3333	0	0	0	1	Covered
0.1429	0.1322	0.2727	0	0	0	1.25	Covered
0.1667	0.1600	0.2000	0	0	0	1.5	Covered
0.2000	0.1975	0.1111	0	0	0	1.75	Covered
0.1250	0.1111	0.3333	0	0.25	0	1	Covered
0.1429	0.1322	0.2727	0	0.25	0	1.25	Covered
0.1667	0.1600	0.2000	0	0.25	0	1.5	Covered
0.2000	0.1975	0.1111	0	0.25	0	1.75	Covered
0.1250	0.1111	0.3333	0	0.5	0	1	Covered
0.1429	0.1322	0.2727	0	0.5	0	1.25	Covered
0.1667	0.1600	0.2000	0	0.5	0	1.5	Covered
0.2000	0.1975	0.1111	0	0.5	0	1.75	Covered
0.1250	0.1111	0.3333	0	0.75	0	1	Covered
0.1429	0.1322	0.2727	0	0.75	0	1.25	Covered
0.1667	0.1600	0.2000	0	0.75	0	1.5	Covered
0.2000	0.1975	0.1111	0	0.75	0	1.75	Covered
0.1250	0.1111	0.3333	0	1	0	1	Covered
0.1429	0.1322	0.2727	0	1	0	1.25	Covered
0.1667	0.1600	0.2000	0	1	0	1.5	Covered
0.2000	0.1975	0.1111	0	1	0	1.75	Covered
0.1250	0.1111	0.3333	0.25	1	0	1	Covered
0.1429	0.1322	0.2727	0.25	1	0	1.25	Covered
0.1667	0.1600	0.2000	0.25	1	0	1.5	Covered
0.2000	0.1975	0.1111	0.25	1	0	1.75	Covered
0.1250	0.1111	0.3333	0.5	1	0	1	Covered
0.1429	0.1322	0.2727	0.5	1	0	1.25	Covered
0.1667	0.1600	0.2000	0.5	1	0	1.5	Covered
0.2000	0.1975	0.1111	0.5	1	0	1.75	Covered

Continued on next page

Table B.1 – continued from previous page

π_c	π_u	d_u	α	β	e	$\bar{\theta}$	Configuration
0.1250	0.1111	0.3333	0.75	1	0	1	Covered
0.1429	0.1322	0.2727	0.75	1	0	1.25	Covered
0.1667	0.1600	0.2000	0.75	1	0	1.5	Covered
0.2000	0.1975	0.1111	0.75	1	0	1.75	Covered
0.1250	0.1111	0.3333	1	1	0	1	Covered
0.1429	0.1322	0.2727	1	1	0	1.25	Covered
0.1667	0.1600	0.2000	1	1	0	1.5	Covered
0.2000	0.1975	0.1111	1	1	0	1.75	Covered
0.1172	0.1085	0.3636	0.25	0.75	0.0625	1	Covered
0.1352	0.1291	0.3000	0.25	0.75	0.0625	1.25	Covered
0.1598	0.1566	0.2222	0.25	0.75	0.0625	1.5	Covered
0.1953	0.1948	0.1250	0.25	0.75	0.0625	1.75	Covered
0.1094	0.1050	0.4000	0.25	0.5	0.125	1	Covered
0.1276	0.1250	0.3333	0.25	0.5	0.125	1.25	Covered
0.1531	0.1523	0.2500	0.25	0.5	0.125	1.5	Covered
0.1914	0.1916	0.0685	0.25	0.5	0.125	1.75	Uncovered
0.1094	0.1050	0.4000	0.5	0.75	0.125	1	Covered
0.1276	0.1250	0.3333	0.5	0.75	0.125	1.25	Covered
0.1531	0.1523	0.2500	0.5	0.75	0.125	1.5	Covered
0.1914	0.1916	0.0685	0.5	0.75	0.125	1.75	Uncovered
0.1016	0.1003	0.4444	0.25	0.25	0.1875	1	Covered
0.1200	0.1196	0.3750	0.25	0.25	0.1875	1.25	Covered
0.1467	0.1467	0.2657	0.25	0.25	0.1875	1.5	Uncovered
0.1886	0.1914	0.0000	0.25	0.25	0.1875	1.75	Uncovered
0.1016	0.1003	0.4444	0.75	0.75	0.1875	1	Covered
0.1200	0.1196	0.3750	0.75	0.75	0.1875	1.25	Covered
0.1467	0.1467	0.2657	0.75	0.75	0.1875	1.5	Uncovered
0.1886	0.1914	0.0000	0.75	0.75	0.1875	1.75	Uncovered
0.0938	0.0938	0.5000	0.25	0	0.25	1	Covered
0.1125	0.1127	0.3731	0.25	0	0.25	1.25	Uncovered
0.1406	0.1425	0.1492	0.25	0	0.25	1.5	Uncovered
0.1875	0.1914	0.0000	0.25	0	0.25	1.75	Uncovered
0.0938	0.0938	0.5000	0.5	0.5	0.25	1	Covered
0.1125	0.1127	0.3731	0.5	0.5	0.25	1.25	Uncovered
0.1406	0.1425	0.1492	0.5	0.5	0.25	1.5	Uncovered
0.1875	0.1914	0.0000	0.5	0.5	0.25	1.75	Uncovered

Continued on next page

Table B.1 – continued from previous page

π_c	π_u	d_u	α	β	e	$\bar{\theta}$	Configuration
0.0938	0.0938	0.5000	1	0.75	0.25	1	Covered
0.1125	0.1127	0.3731	1	0.75	0.25	1.25	Uncovered
0.1406	0.1425	0.1492	1	0.75	0.25	1.5	Uncovered
0.1875	0.1914	0.0000	1	0.75	0.25	1.75	Uncovered
0.0781	0.0817	0.4105	0.5	0.25	0.375	1	Uncovered
0.0977	0.1033	0.2396	0.5	0.25	0.375	1.25	Uncovered
0.1302	0.1406	0.0000	0.5	0.25	0.375	1.5	Uncovered
0.1875	0.1914	0.0000	0.5	0.25	0.375	1.75	Uncovered
0.0781	0.0817	0.4105	0.75	0.5	0.375	1	Uncovered
0.0977	0.1033	0.2396	0.75	0.5	0.375	1.25	Uncovered
0.1302	0.1406	0.0000	0.75	0.5	0.375	1.5	Uncovered
0.1875	0.1914	0.0000	0.75	0.5	0.375	1.75	Uncovered
0.0625	0.0741	0.3333	0.5	0	0.5	1	Uncovered
0.0833	0.0988	0.1111	0.5	0	0.5	1.25	Uncovered
0.1250	0.1406	0.0000	0.5	0	0.5	1.5	Uncovered
0.1875	0.1914	0.0000	0.5	0	0.5	1.75	Uncovered
0.0625	0.0741	0.3333	1	0.5	0.5	1	Uncovered
0.0833	0.0988	0.1111	1	0.5	0.5	1.25	Uncovered
0.1250	0.1406	0.0000	1	0.5	0.5	1.5	Uncovered
0.1875	0.1914	0.0000	1	0.5	0.5	1.75	Uncovered
0.0547	0.0713	0.2967	0.75	0.25	0.5625	1	Uncovered
0.0766	0.0978	0.0433	0.75	0.25	0.5625	1.25	Uncovered
0.1250	0.1406	0.0000	0.75	0.25	0.5625	1.5	Uncovered
0.1875	0.1914	0.0000	0.75	0.25	0.5625	1.75	Uncovered
0.0313	0.0655	0.1861	0.75	0	0.75	1	Uncovered
0.0625	0.0977	0.0000	0.75	0	0.75	1.25	Uncovered
0.1250	0.1406	0.0000	0.75	0	0.75	1.5	Uncovered
0.1875	0.1914	0.0000	0.75	0	0.75	1.75	Uncovered
0.0313	0.0655	0.1861	1	0.25	0.75	1	Uncovered
0.0625	0.0977	0.0000	1	0.25	0.75	1.25	Uncovered
0.1250	0.1406	0.0000	1	0.25	0.75	1.5	Uncovered
0.1875	0.1914	0.0000	1	0.25	0.75	1.75	Uncovered
0.0000	0.0625	0.0000	1	0	1	1	Uncovered
0.0625	0.0977	0.0000	1	0	1	1.25	Uncovered
0.1250	0.1406	0.0000	1	0	1	1.5	Uncovered
0.1875	0.1914	0.0000	1	0	1	1.75	Uncovered

```
Clear["Global`*"]
(* Define model fundamentals *)
SetDirectory[NotebookDirectory[]]
pif[d_, e_, theta_] :=
Piecewise[{{((-1 + d) (-d (-2 + theta) + theta)^2)/(16 (-1 + d -
    2 e d)), (1 <= theta <
    2 && ((0 <= e <= 1/2 &&
    0 <= d < (-2 + theta)/(-4 + 4 e + theta)) || (1/2 < e <=
    1 && 0 <= d < 1))) || (theta >= 2 &&
    1/2 < e <= 1 && (-2 + theta)/(-4 + 4 e + theta) < d < 1)}, {0,
    True}}]
pic[theta_, e_] :=
Piecewise[{{((1 - e)^2)/(4 (3 - theta) - 8 e),
    theta + e <= 2 && theta != 3 - 2 e}, {(theta - 1)/4, True}}]

(* note that constraints for the existence of the solution are already
    within the function *)

(* Optimization *)
result = Table[{N@pic[theta, alpha (1 - beta)],
  N@Maximize[{pif[d, alpha (1 - beta), theta], 0 <= d <= 1}, {d},
  WorkingPrecision -> 4], N@alpha, N@beta, N@theta}, {alpha, 0, 1,
  1/4}, {beta, 0, 1, 1/4}, {theta, 1, 99/100, 1/3}];
flatResult = Flatten /@ Flatten[result, 2];
cleanResult = flatResult /. rule_Rule :> rule /. (d -> x_) :> x
Export["data_uncovered.csv", cleanResult]
```

B.3 First best uncovered market with unaware consumers

We define the indifferent consumer and the demand function as in (B.1), however the welfare function now takes into account the non-user disutility that comes from the

externality and is weighted by the terms $\beta \in (0, 1)$.

$$W_u = \int_{\theta^*}^{\bar{\theta}} \left(\theta \frac{(1-d)^2}{4} - \alpha d \frac{1-d}{2} x - P \right) d\theta - \beta \int_{\bar{\theta}-1}^{\theta^*} \left(\alpha\, x\, d\, \frac{1-d}{2} \right) d\theta + \pi, \tag{B.12}$$

$$\begin{aligned} W_u = & \frac{(4\alpha d + d - 1)\left((d-1)^2\bar{\theta} - 4P\right)^2}{8(d-1)^3} \\ & - \frac{1}{2}\alpha\beta(1-d)d\left(\frac{4P}{(d-1)^2} - \bar{\theta} + 1\right)\left(\bar{\theta} - \frac{4P}{(d-1)^2}\right) \\ & + \left(P - \frac{1}{2}(d-1)d\right)\left(\bar{\theta} - \frac{4P}{(d-1)^2}\right), \end{aligned} \tag{B.13}$$

and the maximization problem would be:

$$\begin{aligned} \max_{d,P} \{W_u\} \quad \text{s.t.} \quad & \frac{4P}{(1-d)^2} > \bar{\theta} - 1, \\ & \frac{4P}{(1-d)^2} \leq \bar{\theta}, \\ & 0 \leq d \leq 1, \\ & 0 \leq \alpha, \beta \leq 1, \end{aligned} \tag{B.14}$$

Differently from the covered market case, this function will depend on prices so we cannot set any price to keep the market uncovered.

Numerical optimization through scipy with the Sequential Least Squares Programming (SLSQP) algorithm highlights that welfare with the uncovered market configuration is always smaller than with the covered market one, and no matter the values of the parameters $\alpha, \beta, \bar{\theta}$ the indifferent consumer is always located at the boundary of the first constraint in (B.14).[1]

Please notice that because of the 4 digits approximation the values of the W_c are equal to W_u. The function W_u converges to W_c without ever reaching it.

Table B.2: Simulation results for planner solution with aware consumers

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
8 0.1667	0.0000	0.3333	0	0	1	0.1667	Covered
0.1923	0.0296	0.2308	0	0	1.2	0.1923	Covered
0.2273	0.0826	0.0909	0	0	1.4	0.2273	Covered

Continued on next page

[1]Full .csv results of the simulations are available at the Drive Folder: https://drive.google.com/drive/folders/1u9XV2XcQLgoAUtsy1N1VPA5MX-i5T9NE?usp=share_link.

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.2750	0.1500	0.0000	0	0	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0	0	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0	0	2	0.3750	Covered
0.4250	0.3000	0.0000	0	0	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0	0	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0	0	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0	0	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0	0	3	0.6250	Covered
0.1667	0.0000	0.3333	0	0.2	1	0.1667	Covered
0.1923	0.0296	0.2308	0	0.2	1.2	0.1923	Covered
0.2273	0.0826	0.0909	0	0.2	1.4	0.2273	Covered
0.2750	0.1500	0.0000	0	0.2	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0	0.2	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0	0.2	2	0.3750	Covered
0.4250	0.3000	0.0000	0	0.2	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0	0.2	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0	0.2	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0	0.2	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0	0.2	3	0.6250	Covered
0.1667	0.0000	0.3333	0	0.4	1	0.1667	Covered
0.1923	0.0296	0.2308	0	0.4	1.2	0.1923	Covered
0.2273	0.0826	0.0909	0	0.4	1.4	0.2273	Covered
0.2750	0.1500	0.0000	0	0.4	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0	0.4	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0	0.4	2	0.3750	Covered
0.4250	0.3000	0.0000	0	0.4	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0	0.4	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0	0.4	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0	0.4	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0	0.4	3	0.6250	Covered
0.1667	0.0000	0.3333	0	0.6	1	0.1667	Covered
0.1923	0.0296	0.2308	0	0.6	1.2	0.1923	Covered
0.2273	0.0826	0.0909	0	0.6	1.4	0.2273	Covered
0.2750	0.1500	0.0000	0	0.6	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0	0.6	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0	0.6	2	0.3750	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.4250	0.3000	0.0000	0	0.6	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0	0.6	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0	0.6	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0	0.6	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0	0.6	3	0.6250	Covered
0.1667	0.0000	0.3333	0	0.8	1	0.1667	Covered
0.1923	0.0296	0.2308	0	0.8	1.2	0.1923	Covered
0.2273	0.0826	0.0909	0	0.8	1.4	0.2273	Covered
0.2750	0.1500	0.0000	0	0.8	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0	0.8	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0	0.8	2	0.3750	Covered
0.4250	0.3000	0.0000	0	0.8	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0	0.8	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0	0.8	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0	0.8	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0	0.8	3	0.6250	Covered
0.1667	0.0000	0.3333	0	1	1	0.1667	Covered
0.1923	0.0296	0.2308	0	1	1.2	0.1923	Covered
0.2273	0.0826	0.0909	0	1	1.4	0.2273	Covered
0.2750	0.1500	0.0000	0	1	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0	1	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0	1	2	0.3750	Covered
0.4250	0.3000	0.0000	0	1	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0	1	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0	1	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0	1	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0	1	3	0.6250	Covered
0.1455	0.0000	0.2727	0.2	0	1	0.1455	Covered
0.1778	0.0395	0.1111	0.2	0	1.2	0.1778	Covered
0.2250	0.1000	0.0000	0.2	0	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.2	0	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.2	0	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.2	0	2	0.3750	Covered
0.4250	0.3000	0.0000	0.2	0	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.2	0	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.2	0	2.6	0.5250	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.5750	0.4500	0.0000	0.2	0	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.2	0	3	0.6250	Covered
0.1455	0.0000	0.2727	0.2	0.2	1	0.1455	Covered
0.1778	0.0395	0.1111	0.2	0.2	1.2	0.1778	Covered
0.2250	0.1000	0.0000	0.2	0.2	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.2	0.2	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.2	0.2	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.2	0.2	2	0.3750	Covered
0.4250	0.3000	0.0000	0.2	0.2	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.2	0.2	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.2	0.2	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.2	0.2	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.2	0.2	3	0.6250	Covered
0.1455	0.0000	0.2727	0.2	0.4	1	0.1455	Covered
0.1778	0.0395	0.1111	0.2	0.4	1.2	0.1778	Covered
0.2250	0.1000	0.0000	0.2	0.4	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.2	0.4	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.2	0.4	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.2	0.4	2	0.3750	Covered
0.4250	0.3000	0.0000	0.2	0.4	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.2	0.4	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.2	0.4	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.2	0.4	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.2	0.4	3	0.6250	Covered
0.1455	0.0000	0.2727	0.2	0.6	1	0.1455	Covered
0.1778	0.0395	0.1111	0.2	0.6	1.2	0.1778	Covered
0.2250	0.1000	0.0000	0.2	0.6	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.2	0.6	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.2	0.6	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.2	0.6	2	0.3750	Covered
0.4250	0.3000	0.0000	0.2	0.6	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.2	0.6	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.2	0.6	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.2	0.6	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.2	0.6	3	0.6250	Covered
0.1455	0.0000	0.2727	0.2	0.8	1	0.1455	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.1778	0.0395	0.1111	0.2	0.8	1.2	0.1778	Covered
0.2250	0.1000	0.0000	0.2	0.8	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.2	0.8	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.2	0.8	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.2	0.8	2	0.3750	Covered
0.4250	0.3000	0.0000	0.2	0.8	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.2	0.8	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.2	0.8	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.2	0.8	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.2	0.8	3	0.6250	Covered
0.1455	0.0000	0.2727	0.2	1	1	0.1455	Covered
0.1778	0.0395	0.1111	0.2	1	1.2	0.1778	Covered
0.2250	0.1000	0.0000	0.2	1	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.2	1	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.2	1	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.2	1	2	0.3750	Covered
0.4250	0.3000	0.0000	0.2	1	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.2	1	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.2	1	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.2	1	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.2	1	3	0.6250	Covered
0.1286	0.0000	0.1429	0.4	0	1	0.1286	Covered
0.1750	0.0500	0.0000	0.4	0	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.4	0	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.4	0	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.4	0	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.4	0	2	0.3750	Covered
0.4250	0.3000	0.0000	0.4	0	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.4	0	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.4	0	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.4	0	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.4	0	3	0.6250	Covered
0.1286	0.0000	0.1429	0.4	0.2	1	0.1286	Covered
0.1750	0.0500	0.0000	0.4	0.2	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.4	0.2	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.4	0.2	1.6	0.2750	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.3250	0.2000	0.0000	0.4	0.2	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.4	0.2	2	0.3750	Covered
0.4250	0.3000	0.0000	0.4	0.2	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.4	0.2	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.4	0.2	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.4	0.2	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.4	0.2	3	0.6250	Covered
0.1286	0.0000	0.1429	0.4	0.4	1	0.1286	Covered
0.1750	0.0500	0.0000	0.4	0.4	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.4	0.4	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.4	0.4	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.4	0.4	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.4	0.4	2	0.3750	Covered
0.4250	0.3000	0.0000	0.4	0.4	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.4	0.4	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.4	0.4	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.4	0.4	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.4	0.4	3	0.6250	Covered
0.1286	0.0000	0.1429	0.4	0.6	1	0.1286	Covered
0.1750	0.0500	0.0000	0.4	0.6	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.4	0.6	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.4	0.6	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.4	0.6	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.4	0.6	2	0.3750	Covered
0.4250	0.3000	0.0000	0.4	0.6	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.4	0.6	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.4	0.6	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.4	0.6	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.4	0.6	3	0.6250	Covered
0.1286	0.0000	0.1429	0.4	0.8	1	0.1286	Covered
0.1750	0.0500	0.0000	0.4	0.8	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.4	0.8	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.4	0.8	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.4	0.8	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.4	0.8	2	0.3750	Covered
0.4250	0.3000	0.0000	0.4	0.8	2.2	0.4250	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.4750	0.3500	0.0000	0.4	0.8	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.4	0.8	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.4	0.8	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.4	0.8	3	0.6250	Covered
0.1286	0.0000	0.1429	0.4	1	1	0.1286	Covered
0.1750	0.0500	0.0000	0.4	1	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.4	1	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.4	1	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.4	1	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.4	1	2	0.3750	Covered
0.4250	0.3000	0.0000	0.4	1	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.4	1	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.4	1	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.4	1	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.4	1	3	0.6250	Covered
0.1250	0.0000	0.0000	0.6	0	1	0.1250	Covered
0.1750	0.0500	0.0000	0.6	0	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.6	0	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.6	0	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.6	0	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.6	0	2	0.3750	Covered
0.4250	0.3000	0.0000	0.6	0	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.6	0	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.6	0	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.6	0	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.6	0	3	0.6250	Covered
0.1250	0.0000	0.0000	0.6	0.2	1	0.1250	Covered
0.1750	0.0500	0.0000	0.6	0.2	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.6	0.2	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.6	0.2	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.6	0.2	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.6	0.2	2	0.3750	Covered
0.4250	0.3000	0.0000	0.6	0.2	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.6	0.2	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.6	0.2	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.6	0.2	2.8	0.5750	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.6250	0.5000	0.0000	0.6	0.2	3	0.6250	Covered
0.1250	0.0000	0.0000	0.6	0.4	1	0.1250	Covered
0.1750	0.0500	0.0000	0.6	0.4	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.6	0.4	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.6	0.4	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.6	0.4	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.6	0.4	2	0.3750	Covered
0.4250	0.3000	0.0000	0.6	0.4	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.6	0.4	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.6	0.4	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.6	0.4	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.6	0.4	3	0.6250	Covered
0.1250	0.0000	0.0000	0.6	0.6	1	0.1250	Covered
0.1750	0.0500	0.0000	0.6	0.6	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.6	0.6	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.6	0.6	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.6	0.6	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.6	0.6	2	0.3750	Covered
0.4250	0.3000	0.0000	0.6	0.6	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.6	0.6	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.6	0.6	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.6	0.6	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.6	0.6	3	0.6250	Covered
0.1250	0.0000	0.0000	0.6	0.8	1	0.1250	Covered
0.1750	0.0500	0.0000	0.6	0.8	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.6	0.8	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.6	0.8	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.6	0.8	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.6	0.8	2	0.3750	Covered
0.4250	0.3000	0.0000	0.6	0.8	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.6	0.8	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.6	0.8	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.6	0.8	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.6	0.8	3	0.6250	Covered
0.1250	0.0000	0.0000	0.6	1	1	0.1250	Covered
0.1750	0.0500	0.0000	0.6	1	1.2	0.1750	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.2250	0.1000	0.0000	0.6	1	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.6	1	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.6	1	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.6	1	2	0.3750	Covered
0.4250	0.3000	0.0000	0.6	1	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.6	1	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.6	1	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.6	1	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.6	1	3	0.6250	Covered
0.1250	0.0000	0.0000	0.8	0	1	0.1250	Covered
0.1750	0.0500	0.0000	0.8	0	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.8	0	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.8	0	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.8	0	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.8	0	2	0.3750	Covered
0.4250	0.3000	0.0000	0.8	0	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.8	0	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.8	0	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.8	0	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.8	0	3	0.6250	Covered
0.1250	0.0000	0.0000	0.8	0.2	1	0.1250	Covered
0.1750	0.0500	0.0000	0.8	0.2	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.8	0.2	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.8	0.2	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.8	0.2	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.8	0.2	2	0.3750	Covered
0.4250	0.3000	0.0000	0.8	0.2	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.8	0.2	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.8	0.2	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.8	0.2	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.8	0.2	3	0.6250	Covered
0.1250	0.0000	0.0000	0.8	0.4	1	0.1250	Covered
0.1750	0.0500	0.0000	0.8	0.4	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.8	0.4	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.8	0.4	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.8	0.4	1.8	0.3250	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.3750	0.2500	0.0000	0.8	0.4	2	0.3750	Covered
0.4250	0.3000	0.0000	0.8	0.4	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.8	0.4	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.8	0.4	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.8	0.4	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.8	0.4	3	0.6250	Covered
0.1250	0.0000	0.0000	0.8	0.6	1	0.1250	Covered
0.1750	0.0500	0.0000	0.8	0.6	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.8	0.6	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.8	0.6	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.8	0.6	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.8	0.6	2	0.3750	Covered
0.4250	0.3000	0.0000	0.8	0.6	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.8	0.6	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.8	0.6	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.8	0.6	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.8	0.6	3	0.6250	Covered
0.1250	0.0000	0.0000	0.8	0.8	1	0.1250	Covered
0.1750	0.0500	0.0000	0.8	0.8	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.8	0.8	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.8	0.8	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.8	0.8	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.8	0.8	2	0.3750	Covered
0.4250	0.3000	0.0000	0.8	0.8	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.8	0.8	2.4	0.4750	Covered
0.5250	0.4000	0.0000	0.8	0.8	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.8	0.8	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.8	0.8	3	0.6250	Covered
0.1250	0.0000	0.0000	0.8	1	1	0.1250	Covered
0.1750	0.0500	0.0000	0.8	1	1.2	0.1750	Covered
0.2250	0.1000	0.0000	0.8	1	1.4	0.2250	Covered
0.2750	0.1500	0.0000	0.8	1	1.6	0.2750	Covered
0.3250	0.2000	0.0000	0.8	1	1.8	0.3250	Covered
0.3750	0.2500	0.0000	0.8	1	2	0.3750	Covered
0.4250	0.3000	0.0000	0.8	1	2.2	0.4250	Covered
0.4750	0.3500	0.0000	0.8	1	2.4	0.4750	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.5250	0.4000	0.0000	0.8	1	2.6	0.5250	Covered
0.5750	0.4500	0.0000	0.8	1	2.8	0.5750	Covered
0.6250	0.5000	0.0000	0.8	1	3	0.6250	Covered
0.1250	0.0000	0.0000	1	0	1	0.1250	Covered
0.1750	0.0500	0.0000	1	0	1.2	0.1750	Covered
0.2250	0.1000	0.0000	1	0	1.4	0.2250	Covered
0.2750	0.1500	0.0000	1	0	1.6	0.2750	Covered
0.3250	0.2000	0.0000	1	0	1.8	0.3250	Covered
0.3750	0.2500	0.0000	1	0	2	0.3750	Covered
0.4250	0.3000	0.0000	1	0	2.2	0.4250	Covered
0.4750	0.3500	0.0000	1	0	2.4	0.4750	Covered
0.5250	0.4000	0.0000	1	0	2.6	0.5250	Covered
0.5750	0.4500	0.0000	1	0	2.8	0.5750	Covered
0.6250	0.5000	0.0000	1	0	3	0.6250	Covered
0.1250	0.0000	0.0000	1	0.2	1	0.1250	Covered
0.1750	0.0500	0.0000	1	0.2	1.2	0.1750	Covered
0.2250	0.1000	0.0000	1	0.2	1.4	0.2250	Covered
0.2750	0.1500	0.0000	1	0.2	1.6	0.2750	Covered
0.3250	0.2000	0.0000	1	0.2	1.8	0.3250	Covered
0.3750	0.2500	0.0000	1	0.2	2	0.3750	Covered
0.4250	0.3000	0.0000	1	0.2	2.2	0.4250	Covered
0.4750	0.3500	0.0000	1	0.2	2.4	0.4750	Covered
0.5250	0.4000	0.0000	1	0.2	2.6	0.5250	Covered
0.5750	0.4500	0.0000	1	0.2	2.8	0.5750	Covered
0.6250	0.5000	0.0000	1	0.2	3	0.6250	Covered
0.1250	0.0000	0.0000	1	0.4	1	0.1250	Covered
0.1750	0.0500	0.0000	1	0.4	1.2	0.1750	Covered
0.2250	0.1000	0.0000	1	0.4	1.4	0.2250	Covered
0.2750	0.1500	0.0000	1	0.4	1.6	0.2750	Covered
0.3250	0.2000	0.0000	1	0.4	1.8	0.3250	Covered
0.3750	0.2500	0.0000	1	0.4	2	0.3750	Covered
0.4250	0.3000	0.0000	1	0.4	2.2	0.4250	Covered
0.4750	0.3500	0.0000	1	0.4	2.4	0.4750	Covered
0.5250	0.4000	0.0000	1	0.4	2.6	0.5250	Covered
0.5750	0.4500	0.0000	1	0.4	2.8	0.5750	Covered
0.6250	0.5000	0.0000	1	0.4	3	0.6250	Covered

Continued on next page

Table B.2 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.1250	0.0000	0.0000	1	0.6	1	0.1250	Covered
0.1750	0.0500	0.0000	1	0.6	1.2	0.1750	Covered
0.2250	0.1000	0.0000	1	0.6	1.4	0.2250	Covered
0.2750	0.1500	0.0000	1	0.6	1.6	0.2750	Covered
0.3250	0.2000	0.0000	1	0.6	1.8	0.3250	Covered
0.3750	0.2500	0.0000	1	0.6	2	0.3750	Covered
0.4250	0.3000	0.0000	1	0.6	2.2	0.4250	Covered
0.4750	0.3500	0.0000	1	0.6	2.4	0.4750	Covered
0.5250	0.4000	0.0000	1	0.6	2.6	0.5250	Covered
0.5750	0.4500	0.0000	1	0.6	2.8	0.5750	Covered
0.6250	0.5000	0.0000	1	0.6	3	0.6250	Covered
0.1250	0.0000	0.0000	1	0.8	1	0.1250	Covered
0.1750	0.0500	0.0000	1	0.8	1.2	0.1750	Covered
0.2250	0.1000	0.0000	1	0.8	1.4	0.2250	Covered
0.2750	0.1500	0.0000	1	0.8	1.6	0.2750	Covered
0.3250	0.2000	0.0000	1	0.8	1.8	0.3250	Covered
0.3750	0.2500	0.0000	1	0.8	2	0.3750	Covered
0.4250	0.3000	0.0000	1	0.8	2.2	0.4250	Covered
0.4750	0.3500	0.0000	1	0.8	2.4	0.4750	Covered
0.5250	0.4000	0.0000	1	0.8	2.6	0.5250	Covered
0.5750	0.4500	0.0000	1	0.8	2.8	0.5750	Covered
0.6250	0.5000	0.0000	1	0.8	3	0.6250	Covered
0.1250	0.0000	0.0000	1	1	1	0.1250	Covered
0.1750	0.0500	0.0000	1	1	1.2	0.1750	Covered
0.2250	0.1000	0.0000	1	1	1.4	0.2250	Covered
0.2750	0.1500	0.0000	1	1	1.6	0.2750	Covered
0.3250	0.2000	0.0000	1	1	1.8	0.3250	Covered
0.3750	0.2500	0.0000	1	1	2	0.3750	Covered
0.4250	0.3000	0.0000	1	1	2.2	0.4250	Covered
0.4750	0.3500	0.0000	1	1	2.4	0.4750	Covered
0.5250	0.4000	0.0000	1	1	2.6	0.5250	Covered
0.5750	0.4500	0.0000	1	1	2.8	0.5750	Covered
0.6250	0.5000	0.0000	1	1	3	0.6250	Covered

Mathematica Code Here is reported the Mathematica code used for the simulation.[2]

```
Clear["Global`*"]
(* Define model fundamentals *)

tsc = Simplify[ 4 P/(1 \[Minus] d)^2];
THB = THU - 1;
xu = Simplify[ THU - tsc];
uiu = Simplify[( theta (1 - d)^2)/4 -
alpha (THU - tsc) d (1 - d)/2 - P];
u0u = Simplify[-alpha beta (THU - tsc) d (1 - d)/2];
profit = Simplify[(THU - tsc) (P + (d - d^2)/2)];
uw = \!\(
\*SubsuperscriptBox[\(\[Integral]\), \(tsc\), \(THU\)]\(\((uiu + \((P \
+
\*FractionBox[\((d -
\*SuperscriptBox[\(d\), \(2\)])\), \(2\)])\))\) \[DifferentialD]theta\
\)\) + \!\(
\*SubsuperscriptBox[\(\[Integral]\), \(THU -
1\), \(tsc\)]\(u0u \[DifferentialD]theta\)\);

(* welfare function *)
cwf[alpha_, THU_] :=
Piecewise[{{-((-1 + alpha)^2/(-10 + 8*alpha + 4*THU)),
    1/2 < THU < 3/2 && alpha >= 0 && alpha + THU <= 3/2}}, (-1 +
2*THU)/8]
welfun[P_, d_, beta_, alpha_,
THU_ ] := -(1/(
8 (-1 + d)^3)) (-4 P + (-1 + d)^2 THU) (4 P + THU +
d^3 (4 + 4 alpha beta (-1 + THU) - THU - 4 alpha THU) +
d^2 (-8 + 3 THU + 8 alpha (beta + THU - beta THU)) -
d (-4 + 4 P + 3 THU +
4 alpha (beta - 4 P + 4 beta P + THU - beta THU)))


(* run simulation *)
result =
Table[{N@Maximize[{welfun[P, d, beta, alpha, THU],
    0 <= d <= 1, (4 P)/(1 - 2 d + d^2) <= THU <= (
    1 - 2 d + d^2 + 4 P)/(1 - 2 d + d^2)}, {P, d},
  WorkingPrecision -> 6], N@alpha, N@beta, N@THU,
```

[2]Scipy's python optimization confirms this results and is available upon request.

```
    N@cwf[alpha, THU]}, {alpha, 0, 1, 1/5}, {beta, 0, 1, 1/5}, {THU, 1,
    3, 1/5}]

(* clean results *)
flatResult = Flatten /@ Flatten[result, 2];
cleanResult = flatResult /. rule_Rule :> rule[[2]];

SetDirectory["InsertYourPath"]
Export["data_uncovered_welfare_unaw.csv", cleanResult]
```

B.4 First best with aware consumers

Covered market with aware consumers When consumers can internalize the externality and the market is covered the indifferent consumer equation is (B.1) given that prices do not influence welfare when the market is covered P is set to ensure a covered market.

$$P = \frac{1}{4}(1-d)(2\alpha(\beta-1)d - d\bar{\theta} + d + \bar{\theta} - 1),$$

and the welfare function is the sum of CS and profit:

$$W_a^c = \int_{\theta^*}^{\bar{\theta}} U_i d\theta + \pi = \frac{1}{8}(d-1)(d(4\alpha + 2\bar{\theta} - 5) - 2\bar{\theta} + 1),$$

where maximization in d of this function gives:

$$d_{fb}^a = \begin{cases} \dfrac{2\alpha + 2\bar{\theta} - 3}{4\alpha + 2\bar{\theta} - 5} & 1 \leq \bar{\theta} \leq \dfrac{3}{2} \wedge \alpha \leq \dfrac{1}{2}(3 - 2\bar{\theta}), \\ 0 & \text{Otherwise,} \end{cases} \tag{B.15}$$

and total welfare is:

$$W_{fb}^a = \begin{cases} \dfrac{(\alpha-1)^2}{10 - 8\alpha - 4\bar{\theta}} & 1 \leq \bar{\theta} \leq \dfrac{3}{2} \wedge \alpha + \bar{\theta} \leq \dfrac{3}{2}, \\ \dfrac{1}{8}(2\bar{\theta} - 1) & \text{Otherwise,} \end{cases} \tag{B.16}$$

Prices are just a transfer from consumers to the firm and it turns out that the price that keeps the market covered is:

$$P_{fb}^a = \begin{cases} \dfrac{(\alpha-1)\left(2\alpha^2(\beta-1) + \alpha(\beta(2\bar{\theta}-3) - \bar{\theta} + 2) - \bar{\theta} + 1\right)}{(4\alpha + 2\bar{\theta} - 5)^2} & 1 \leq \bar{\theta} \leq \dfrac{3}{2} \wedge \alpha + \bar{\theta} \leq \dfrac{3}{2}, \\ \dfrac{\bar{\theta} - 1}{4} & \text{Otherwise,} \end{cases}$$

even if the price that is different from the one found in section 2.3.2, the welfare function remains the same.

The extension with welfare maximization under an uncovered market and aware consumers is treated in B.4

Uncovered market with aware consumers

$$
\begin{aligned}
W_u = &\int_{\theta^*}^{\bar{\theta}} \left(\theta \frac{(1-d)^2}{4} - \alpha d \frac{1-d}{2} x - P \right) \, d\theta \\
&- \int_{\bar{\theta}}^{\theta^*} \left(\alpha \, x \, \beta \, d \, \frac{1-d}{2} \right) \, d\theta + \pi, \qquad \text{(B.17)} \\
= &\frac{1}{8(d-1)(2\alpha(\beta-1)d+d-1)^2} \Big((d-1)^2 \bar{\theta} - 4P \Big) \\
&\Big(d^3 \Big(8\alpha^2(\beta-1)\beta - 4\alpha(\beta-2) + \bar{\theta} - 4 \Big) \\
&+ d^2 \Big(-8\alpha^2(\beta-1)\beta - 8\alpha - 3\bar{\theta} + 8 \Big) \\
&+ d(4\alpha(4\beta P + \beta - 4P) + 4P + 3\bar{\theta} - 4) - 4P - \bar{\theta} \Big),
\end{aligned}
$$

here the price is not only a transfer because a change in price would change the market coverage and modifies welfare. So to maximize this function we should write the Lagrangean for the constrained maximization problem where the constraints are: $\bar{\theta} - 1 < \theta^* \leq \bar{\theta}$ and $0 \leq d \leq 1$.

When expanding the integral and writing the Lagrangian with four constraints we have an overwhelmingly complex problem. Inspection of numerical simulation, shows that the solution lies on the boundary $\bar{\theta} - 1 < \theta^*$, and that the welfare generated by this solution is lower than the one provided in the covered market case. [3]

[3] Full .csv results of the simulations are available at the Drive Folder: https://drive.google.com/drive/folders/1u9XV2XcQLgoAUtsy1N1VPA5MX-i5T9NE?usp=share_link.

Mathematica Code Here is reported the Mathematica code that replicates the simulation.

```
1
2  Clear["Global`*"]
3  (* Define models fundamentals *)
4
5  ui = Simplify[( theta (1 - d)^2)/4 - alpha x d (1 - d)/2 - P];
6  u0 = Simplify[-alpha beta x d (1 - d)/2];
7  x = FullSimplify[ thetabar - theta];
8  Reduce[ui == u0, theta];
9  indcons =
10 Simplify[ (
11 2 (2 P + alpha d thetabar - alpha beta d thetabar -
12 alpha d^2 thetabar + alpha beta d^2 thetabar))/((-1 + d) (-1 +
13 d - 2 alpha d + 2 alpha beta d))];
14 x = Simplify[ thetabar - indcons];
15 uiu = Simplify[( theta (1 - d)^2)/4 - alpha x d (1 - d)/2 - P];
16 u0u = Simplify[-alpha beta x d (1 - d)/2];
17 profit = Simplify[x (P + (d - d^2)/2)];
18 uw = Simplify[\!\(
19 \*SubsuperscriptBox[\(\[Integral]\), \(indcons\), \
20 \(thetabar\)]\(\((uiu + \ \((P +
21 \*FractionBox[\((d -
22 \*SuperscriptBox[\(d\), \(2\)])\), \(2\)])\))\) \[DifferentialD]theta\
23 \)\) + \!\(
24 \*SubsuperscriptBox[\(\[Integral]\), \(thetabar -
25 1\), \(indcons\)]\(u0u \[DifferentialD]theta\)\)];
26
27 (* Define simulation function NB: w = uw *)
28 Clear[alpha, beta, thetabar]
29 cwf[alpha_, thetabar_] :=
30 Piecewise[{{-((-1 + alpha)^2/(-10 + 8 alpha + 4 thetabar)),
31     1 <= thetabar <= 3/2 && alpha + thetabar <= 3/2}},
32 1/8 (-1 + 2 thetabar)]
33 w[P_, beta_, alpha_, thetabar_,
34 d_] = ((-4 P + (-1 + d)^2 thetabar) (-4 P +
35 d^2 (8 - 8 alpha - 8 alpha^2 (-1 + beta) beta - 3 thetabar) -
36 thetabar + d^3 (-4 - 4 alpha (-2 + beta) + 8 alpha^2 (-1 + beta) beta +
37 thetabar) + d (-4 + 4 P + 4 alpha (beta - 4 P + 4 beta P) +
38 3 thetabar)))/(8 (-1 + d) (-1 + d +
39 2 alpha (-1 + beta) d)^2);
```

```
40
41 (* Run Simulation *)
42 result =
43 Table[{NMaximize[{w[P, beta, alpha, thetabar, d], 0 <= d <= 1,
44     thetabar - 1 < indcons <= thetabar}, {P, d},
45   WorkingPrecision -> 6], N@alpha, N@beta, N@thetabar,
46   cwf[alpha, thetabar]}, {alpha, 0, 1, 1/5}, {beta, 0, 1,
47   1/5}, {thetabar, 1, 15/5, 1/5}]
48
49 (*Clean Results*)
50 flatResult=Flatten/@Flatten[result,2]
51 cleanResult=flatResult/.rule_Rule:>rule[[2]]
52
53 (*Set save directory*)
54 SetDirectory["InsertYourPath"];
55
56 (*Export to .csv*)
57 Export["results_uncovered_welfare_aware.csv",cleanResult]
```

Please notice that because of the 4 digits approximation the values of the W_c are equal to W_u. The function W_u converges to W_c without ever reaching it:

Table B.3: Simulation results for monopolist equilibrium with aware consumers

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
8 0.16666	0.00000	0.33746	0	0	1	0.16667	Covered
0.19230	0.02932	0.23428	0	0	1.2	0.19231	Covered
0.22727	0.08204	0.09426	0	0	1.4	0.22727	Covered
0.27477	0.14871	0.00438	0	0	1.6	0.27500	Covered
0.32467	0.19913	0.00219	0	0	1.8	0.32500	Covered
0.37488	0.24977	0.00048	0	0	2	0.37500	Covered
0.42497	0.29996	0.00008	0	0	2.2	0.42500	Covered
0.47485	0.35011	0.00000	0	0	2.4	0.47500	Covered
0.52464	0.40023	0.00000	0	0	2.6	0.52500	Covered
0.57467	0.45019	0.00000	0	0	2.8	0.57500	Covered
0.62415	0.50043	0.00000	0	0	3	0.62500	Covered
0.16666	0.00000	0.33746	0	0.2	1	0.16667	Covered
0.19230	0.02932	0.23428	0	0.2	1.2	0.19231	Covered
0.22727	0.08204	0.09426	0	0.2	1.4	0.22727	Covered
0.27477	0.14871	0.00438	0	0.2	1.6	0.27500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.32467	0.19913	0.00219	0	0.2	1.8	0.32500	Covered
0.37488	0.24977	0.00048	0	0.2	2	0.37500	Covered
0.42497	0.29996	0.00008	0	0.2	2.2	0.42500	Covered
0.47485	0.35011	0.00000	0	0.2	2.4	0.47500	Covered
0.52464	0.40023	0.00000	0	0.2	2.6	0.52500	Covered
0.57467	0.45019	0.00000	0	0.2	2.8	0.57500	Covered
0.62415	0.50043	0.00000	0	0.2	3	0.62500	Covered
0.16666	0.00000	0.33746	0	0.4	1	0.16667	Covered
0.19230	0.02932	0.23428	0	0.4	1.2	0.19231	Covered
0.22727	0.08204	0.09426	0	0.4	1.4	0.22727	Covered
0.27477	0.14871	0.00438	0	0.4	1.6	0.27500	Covered
0.32467	0.19913	0.00219	0	0.4	1.8	0.32500	Covered
0.37488	0.24977	0.00048	0	0.4	2	0.37500	Covered
0.42497	0.29996	0.00008	0	0.4	2.2	0.42500	Covered
0.47485	0.35011	0.00000	0	0.4	2.4	0.47500	Covered
0.52464	0.40023	0.00000	0	0.4	2.6	0.52500	Covered
0.57467	0.45019	0.00000	0	0.4	2.8	0.57500	Covered
0.62415	0.50043	0.00000	0	0.4	3	0.62500	Covered
0.16666	0.00000	0.33746	0	0.6	1	0.16667	Covered
0.19230	0.02932	0.23428	0	0.6	1.2	0.19231	Covered
0.22727	0.08204	0.09426	0	0.6	1.4	0.22727	Covered
0.27477	0.14871	0.00438	0	0.6	1.6	0.27500	Covered
0.32467	0.19913	0.00219	0	0.6	1.8	0.32500	Covered
0.37488	0.24977	0.00048	0	0.6	2	0.37500	Covered
0.42497	0.29996	0.00008	0	0.6	2.2	0.42500	Covered
0.47485	0.35011	0.00000	0	0.6	2.4	0.47500	Covered
0.52464	0.40023	0.00000	0	0.6	2.6	0.52500	Covered
0.57467	0.45019	0.00000	0	0.6	2.8	0.57500	Covered
0.62415	0.50043	0.00000	0	0.6	3	0.62500	Covered
0.16666	0.00000	0.33746	0	0.8	1	0.16667	Covered
0.19230	0.02932	0.23428	0	0.8	1.2	0.19231	Covered
0.22727	0.08204	0.09426	0	0.8	1.4	0.22727	Covered
0.27477	0.14871	0.00438	0	0.8	1.6	0.27500	Covered
0.32467	0.19913	0.00219	0	0.8	1.8	0.32500	Covered
0.37488	0.24977	0.00048	0	0.8	2	0.37500	Covered
0.42497	0.29996	0.00008	0	0.8	2.2	0.42500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.47485	0.35011	0.00000	0	0.8	2.4	0.47500	Covered
0.52464	0.40023	0.00000	0	0.8	2.6	0.52500	Covered
0.57467	0.45019	0.00000	0	0.8	2.8	0.57500	Covered
0.62415	0.50043	0.00000	0	0.8	3	0.62500	Covered
0.16666	0.00000	0.33746	0	1	1	0.16667	Covered
0.19230	0.02932	0.23428	0	1	1.2	0.19231	Covered
0.22727	0.08204	0.09426	0	1	1.4	0.22727	Covered
0.27477	0.14871	0.00438	0	1	1.6	0.27500	Covered
0.32467	0.19913	0.00219	0	1	1.8	0.32500	Covered
0.37488	0.24977	0.00048	0	1	2	0.37500	Covered
0.42497	0.29996	0.00008	0	1	2.2	0.42500	Covered
0.47485	0.35011	0.00000	0	1	2.4	0.47500	Covered
0.52464	0.40023	0.00000	0	1	2.6	0.52500	Covered
0.57467	0.45019	0.00000	0	1	2.8	0.57500	Covered
0.62415	0.50043	0.00000	0	1	3	0.62500	Covered
0.14545	-0.02004	0.27748	0.2	0	1	0.14545	Covered
0.17777	0.02887	0.11569	0.2	0	1.2	0.17778	Covered
0.22476	0.09861	0.00466	0.2	0	1.4	0.22500	Covered
0.27471	0.14932	0.00178	0.2	0	1.6	0.27500	Covered
0.32489	0.19981	0.00039	0.2	0	1.8	0.32500	Covered
0.37491	0.25009	0.00000	0.2	0	2	0.37500	Covered
0.42487	0.30011	0.00000	0.2	0	2.2	0.42500	Covered
0.47467	0.35024	0.00000	0.2	0	2.4	0.47500	Covered
0.52465	0.40022	0.00000	0.2	0	2.6	0.52500	Covered
0.57377	0.45068	0.00000	0.2	0	2.8	0.57500	Covered
0.62342	0.50079	0.00000	0.2	0	3	0.62500	Covered
0.14541	-0.01626	0.28440	0.2	0.2	1	0.14545	Covered
0.17777	0.03088	0.11604	0.2	0.2	1.2	0.17778	Covered
0.22481	0.09895	0.00377	0.2	0.2	1.4	0.22500	Covered
0.27480	0.14952	0.00130	0.2	0.2	1.6	0.27500	Covered
0.32484	0.19971	0.00062	0.2	0.2	1.8	0.32500	Covered
0.37500	0.25000	0.00000	0.2	0.2	2	0.37500	Covered
0.42492	0.30007	0.00000	0.2	0.2	2.2	0.42500	Covered
0.47466	0.35024	0.00000	0.2	0.2	2.4	0.47500	Covered
0.52462	0.40024	0.00000	0.2	0.2	2.6	0.52500	Covered
0.57408	0.45051	0.00000	0.2	0.2	2.8	0.57500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.62388	0.50056	0.00000	0.2	0.2	3	0.62500	Covered
0.14544	-0.01204	0.27845	0.2	0.4	1	0.14545	Covered
0.17776	0.03248	0.11931	0.2	0.4	1.2	0.17778	Covered
0.22471	0.09854	0.00566	0.2	0.4	1.4	0.22500	Covered
0.27470	0.14930	0.00198	0.2	0.4	1.6	0.27500	Covered
0.32489	0.19981	0.00042	0.2	0.4	1.8	0.32500	Covered
0.37495	0.25006	0.00000	0.2	0.4	2	0.37500	Covered
0.42492	0.30007	0.00000	0.2	0.4	2.2	0.42500	Covered
0.47475	0.35018	0.00000	0.2	0.4	2.4	0.47500	Covered
0.52473	0.40017	0.00000	0.2	0.4	2.6	0.52500	Covered
0.57421	0.45044	0.00000	0.2	0.4	2.8	0.57500	Covered
0.62407	0.50047	0.00000	0.2	0.4	3	0.62500	Covered
0.14542	-0.00811	0.28362	0.2	0.6	1	0.14545	Covered
0.17776	0.03481	0.11758	0.2	0.6	1.2	0.17778	Covered
0.22475	0.09887	0.00481	0.2	0.6	1.4	0.22500	Covered
0.27477	0.14948	0.00154	0.2	0.6	1.6	0.27500	Covered
0.32483	0.19971	0.00068	0.2	0.6	1.8	0.32500	Covered
0.37496	0.24995	0.00009	0.2	0.6	2	0.37500	Covered
0.42491	0.30008	0.00000	0.2	0.6	2.2	0.42500	Covered
0.47478	0.35016	0.00000	0.2	0.6	2.4	0.47500	Covered
0.52444	0.40035	0.00000	0.2	0.6	2.6	0.52500	Covered
0.57394	0.45059	0.00000	0.2	0.6	2.8	0.57500	Covered
0.62391	0.50055	0.00000	0.2	0.6	3	0.62500	Covered
0.14544	-0.00400	0.27738	0.2	0.8	1	0.14545	Covered
0.17777	0.03714	0.11492	0.2	0.8	1.2	0.17778	Covered
0.22483	0.09929	0.00326	0.2	0.8	1.4	0.22500	Covered
0.27460	0.14922	0.00253	0.2	0.8	1.6	0.27500	Covered
0.32480	0.19967	0.00079	0.2	0.8	1.8	0.32500	Covered
0.37491	0.24989	0.00023	0.2	0.8	2	0.37500	Covered
0.42494	0.30005	0.00000	0.2	0.8	2.2	0.42500	Covered
0.47476	0.35017	0.00000	0.2	0.8	2.4	0.47500	Covered
0.52461	0.40024	0.00000	0.2	0.8	2.6	0.52500	Covered
0.57401	0.45055	0.00000	0.2	0.8	2.8	0.57500	Covered
0.62398	0.50051	0.00000	0.2	0.8	3	0.62500	Covered
0.14545	0.00001	0.27444	0.2	1	1	0.14545	Covered
0.17777	0.03901	0.11675	0.2	1	1.2	0.17778	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.22468	0.09877	0.00617	0.2	1	1.4	0.22500	Covered
0.27469	0.14939	0.00205	0.2	1	1.6	0.27500	Covered
0.32477	0.19964	0.00092	0.2	1	1.8	0.32500	Covered
0.37493	0.24997	0.00012	0.2	1	2	0.37500	Covered
0.42496	0.30003	0.00000	0.2	1	2.2	0.42500	Covered
0.47476	0.35017	0.00000	0.2	1	2.4	0.47500	Covered
0.52463	0.40023	0.00000	0.2	1	2.6	0.52500	Covered
0.57427	0.45040	0.00000	0.2	1	2.8	0.57500	Covered
0.62428	0.50036	0.00000	0.2	1	3	0.62500	Covered
0.12857	-0.02447	0.14274	0.4	0	1	0.12857	Covered
0.17490	0.04939	0.00204	0.4	0	1.2	0.17500	Covered
0.22495	0.09987	0.00032	0.4	0	1.4	0.22500	Covered
0.27497	0.14998	0.00008	0.4	0	1.6	0.27500	Covered
0.32489	0.20014	0.00000	0.4	0	1.8	0.32500	Covered
0.37485	0.25015	0.00000	0.4	0	2	0.37500	Covered
0.42465	0.30029	0.00000	0.4	0	2.2	0.42500	Covered
0.47440	0.35043	0.00000	0.4	0	2.4	0.47500	Covered
0.52422	0.40049	0.00000	0.4	0	2.6	0.52500	Covered
0.57405	0.45053	0.00000	0.4	0	2.8	0.57500	Covered
0.62382	0.50059	0.00000	0.4	0	3	0.62500	Covered
0.12837	-0.01556	0.10919	0.4	0.2	1	0.12857	Covered
0.17488	0.04940	0.00234	0.4	0.2	1.2	0.17500	Covered
0.22491	0.09979	0.00060	0.4	0.2	1.4	0.22500	Covered
0.27498	0.14997	0.00008	0.4	0.2	1.6	0.27500	Covered
0.32494	0.20007	0.00000	0.4	0.2	1.8	0.32500	Covered
0.37491	0.25009	0.00000	0.4	0.2	2	0.37500	Covered
0.42482	0.30015	0.00000	0.4	0.2	2.2	0.42500	Covered
0.47476	0.35017	0.00000	0.4	0.2	2.4	0.47500	Covered
0.52436	0.40040	0.00000	0.4	0.2	2.6	0.52500	Covered
0.57391	0.45061	0.00000	0.4	0.2	2.8	0.57500	Covered
0.62374	0.50063	0.00000	0.4	0.2	3	0.62500	Covered
0.12856	-0.01529	0.15000	0.4	0.4	1	0.12857	Covered
0.17431	0.04710	0.01338	0.4	0.4	1.2	0.17500	Covered
0.22484	0.09967	0.00106	0.4	0.4	1.4	0.22500	Covered
0.27487	0.14993	0.00036	0.4	0.4	1.6	0.27500	Covered
0.32499	0.19999	0.00002	0.4	0.4	1.8	0.32500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.37497	0.25003	0.00000	0.4	0.4	2	0.37500	Covered
0.42479	0.30018	0.00000	0.4	0.4	2.2	0.42500	Covered
0.47481	0.35014	0.00000	0.4	0.4	2.4	0.47500	Covered
0.52436	0.40040	0.00000	0.4	0.4	2.6	0.52500	Covered
0.57417	0.45046	0.00000	0.4	0.4	2.8	0.57500	Covered
0.62355	0.50073	0.00000	0.4	0.4	3	0.62500	Covered
0.12857	-0.00980	0.14286	0.4	0.6	1	0.12857	Covered
0.17476	0.04914	0.00480	0.4	0.6	1.2	0.17500	Covered
0.22472	0.09948	0.00188	0.4	0.6	1.4	0.22500	Covered
0.27481	0.14975	0.00070	0.4	0.6	1.6	0.27500	Covered
0.32498	0.19999	0.00004	0.4	0.6	1.8	0.32500	Covered
0.37495	0.25005	0.00000	0.4	0.6	2	0.37500	Covered
0.42482	0.30015	0.00000	0.4	0.6	2.2	0.42500	Covered
0.47484	0.35011	0.00000	0.4	0.6	2.4	0.47500	Covered
0.52453	0.40029	0.00000	0.4	0.6	2.6	0.52500	Covered
0.57430	0.45039	0.00000	0.4	0.6	2.8	0.57500	Covered
0.62400	0.50050	0.00000	0.4	0.6	3	0.62500	Covered
0.12857	-0.00503	0.14774	0.4	0.8	1	0.12857	Covered
0.17481	0.04948	0.00374	0.4	0.8	1.2	0.17500	Covered
0.22486	0.09979	0.00090	0.4	0.8	1.4	0.22500	Covered
0.27479	0.14976	0.00077	0.4	0.8	1.6	0.27500	Covered
0.32491	0.19993	0.00021	0.4	0.8	1.8	0.32500	Covered
0.37498	0.25002	0.00000	0.4	0.8	2	0.37500	Covered
0.42487	0.30011	0.00000	0.4	0.8	2.2	0.42500	Covered
0.47476	0.35017	0.00000	0.4	0.8	2.4	0.47500	Covered
0.52442	0.40036	0.00000	0.4	0.8	2.6	0.52500	Covered
0.57439	0.45034	0.00000	0.4	0.8	2.8	0.57500	Covered
0.62387	0.50057	0.00000	0.4	0.8	3	0.62500	Covered
0.12853	0.00003	0.12856	0.4	1	1	0.12857	Covered
0.17479	0.04958	0.00425	0.4	1	1.2	0.17500	Covered
0.22481	0.09976	0.00124	0.4	1	1.4	0.22500	Covered
0.27478	0.14976	0.00085	0.4	1	1.6	0.27500	Covered
0.32493	0.19993	0.00019	0.4	1	1.8	0.32500	Covered
0.37500	0.25000	0.00000	0.4	1	2	0.37500	Covered
0.42492	0.30007	0.00000	0.4	1	2.2	0.42500	Covered
0.47473	0.35019	0.00000	0.4	1	2.4	0.47500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.52478	0.40014	0.00000	0.4	1	2.6	0.52500	Covered
0.57434	0.45037	0.00000	0.4	1	2.8	0.57500	Covered
0.62401	0.50050	0.00000	0.4	1	3	0.62500	Covered
0.12500	0.00086	0.00000	0.6	0	1	0.12500	Covered
0.17499	0.04997	0.00009	0.6	0	1.2	0.17500	Covered
0.22496	0.10009	0.00000	0.6	0	1.4	0.22500	Covered
0.27493	0.15012	0.00000	0.6	0	1.6	0.27500	Covered
0.32485	0.20019	0.00000	0.6	0	1.8	0.32500	Covered
0.37478	0.25022	0.00000	0.6	0	2	0.37500	Covered
0.42454	0.30038	0.00000	0.6	0	2.2	0.42500	Covered
0.47404	0.35069	0.00000	0.6	0	2.4	0.47500	Covered
0.52391	0.40068	0.00000	0.6	0	2.6	0.52500	Covered
0.57405	0.45053	0.00000	0.6	0	2.8	0.57500	Covered
0.62285	0.50108	0.00000	0.6	0	3	0.62500	Covered
0.12500	0.00144	0.00000	0.6	0.2	1	0.12500	Covered
0.17497	0.04996	0.00019	0.6	0.2	1.2	0.17500	Covered
0.22499	0.10002	0.00000	0.6	0.2	1.4	0.22500	Covered
0.27491	0.15015	0.00000	0.6	0.2	1.6	0.27500	Covered
0.32490	0.20012	0.00000	0.6	0.2	1.8	0.32500	Covered
0.37477	0.25023	0.00000	0.6	0.2	2	0.37500	Covered
0.42467	0.30028	0.00000	0.6	0.2	2.2	0.42500	Covered
0.47429	0.35050	0.00000	0.6	0.2	2.4	0.47500	Covered
0.52412	0.40055	0.00000	0.6	0.2	2.6	0.52500	Covered
0.57375	0.45069	0.00000	0.6	0.2	2.8	0.57500	Covered
0.62383	0.50058	0.00000	0.6	0.2	3	0.62500	Covered
0.12500	0.00059	0.00000	0.6	0.4	1	0.12500	Covered
0.17499	0.04998	0.00006	0.6	0.4	1.2	0.17500	Covered
0.22499	0.09999	0.00003	0.6	0.4	1.4	0.22500	Covered
0.27498	0.15003	0.00000	0.6	0.4	1.6	0.27500	Covered
0.32489	0.20014	0.00000	0.6	0.4	1.8	0.32500	Covered
0.37485	0.25016	0.00000	0.6	0.4	2	0.37500	Covered
0.42473	0.30023	0.00000	0.6	0.4	2.2	0.42500	Covered
0.47435	0.35047	0.00000	0.6	0.4	2.4	0.47500	Covered
0.52434	0.40041	0.00000	0.6	0.4	2.6	0.52500	Covered
0.57389	0.45062	0.00000	0.6	0.4	2.8	0.57500	Covered
0.62382	0.50059	0.00000	0.6	0.4	3	0.62500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.12500	0.00046	0.00000	0.6	0.6	1	0.12500	Covered
0.17496	0.04996	0.00026	0.6	0.6	1.2	0.17500	Covered
0.22493	0.09993	0.00026	0.6	0.6	1.4	0.22500	Covered
0.27495	0.15003	0.00005	0.6	0.6	1.6	0.27500	Covered
0.32491	0.20011	0.00000	0.6	0.6	1.8	0.32500	Covered
0.37492	0.25008	0.00000	0.6	0.6	2	0.37500	Covered
0.42479	0.30018	0.00000	0.6	0.6	2.2	0.42500	Covered
0.47479	0.35015	0.00000	0.6	0.6	2.4	0.47500	Covered
0.52420	0.40050	0.00000	0.6	0.6	2.6	0.52500	Covered
0.57403	0.45054	0.00000	0.6	0.6	2.8	0.57500	Covered
0.62329	0.50086	0.00000	0.6	0.6	3	0.62500	Covered
0.12500	0.00059	0.00000	0.6	0.8	1	0.12500	Covered
0.17493	0.04994	0.00044	0.6	0.8	1.2	0.17500	Covered
0.22489	0.09991	0.00041	0.6	0.8	1.4	0.22500	Covered
0.27490	0.14994	0.00024	0.6	0.8	1.6	0.27500	Covered
0.32497	0.20003	0.00000	0.6	0.8	1.8	0.32500	Covered
0.37493	0.25007	0.00000	0.6	0.8	2	0.37500	Covered
0.42483	0.30014	0.00000	0.6	0.8	2.2	0.42500	Covered
0.47455	0.35032	0.00000	0.6	0.8	2.4	0.47500	Covered
0.52419	0.40051	0.00000	0.6	0.8	2.6	0.52500	Covered
0.57435	0.45036	0.00000	0.6	0.8	2.8	0.57500	Covered
0.62347	0.50077	0.00000	0.6	0.8	3	0.62500	Covered
0.12500	0.00086	0.00000	0.6	1	1	0.12500	Covered
0.17495	0.04999	0.00029	0.6	1	1.2	0.17500	Covered
0.22489	0.09991	0.00046	0.6	1	1.4	0.22500	Covered
0.27490	0.14991	0.00029	0.6	1	1.6	0.27500	Covered
0.32494	0.19996	0.00011	0.6	1	1.8	0.32500	Covered
0.37493	0.25007	0.00000	0.6	1	2	0.37500	Covered
0.42489	0.30009	0.00000	0.6	1	2.2	0.42500	Covered
0.47482	0.35013	0.00000	0.6	1	2.4	0.47500	Covered
0.52472	0.40017	0.00000	0.6	1	2.6	0.52500	Covered
0.57441	0.45033	0.00000	0.6	1	2.8	0.57500	Covered
0.62422	0.50039	0.00000	0.6	1	3	0.62500	Covered
0.12497	0.00403	0.00000	0.8	0	1	0.12500	Covered
0.17498	0.05012	0.00000	0.8	0	1.2	0.17500	Covered
0.22495	0.10013	0.00000	0.8	0	1.4	0.22500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.27489	0.15018	0.00000	0.8	0	1.6	0.27500	Covered
0.32475	0.20031	0.00000	0.8	0	1.8	0.32500	Covered
0.37441	0.25059	0.00000	0.8	0	2	0.37500	Covered
0.42425	0.30063	0.00000	0.8	0	2.2	0.42500	Covered
0.47381	0.35085	0.00000	0.8	0	2.4	0.47500	Covered
0.52349	0.40094	0.00000	0.8	0	2.6	0.52500	Covered
0.57322	0.45099	0.00000	0.8	0	2.8	0.57500	Covered
0.62351	0.50075	0.00000	0.8	0	3	0.62500	Covered
0.12500	0.00118	0.00001	0.8	0.2	1	0.12500	Covered
0.17499	0.05006	0.00000	0.8	0.2	1.2	0.17500	Covered
0.22495	0.10012	0.00000	0.8	0.2	1.4	0.22500	Covered
0.27494	0.15009	0.00000	0.8	0.2	1.6	0.27500	Covered
0.32483	0.20022	0.00000	0.8	0.2	1.8	0.32500	Covered
0.37454	0.25046	0.00000	0.8	0.2	2	0.37500	Covered
0.42423	0.30064	0.00000	0.8	0.2	2.2	0.42500	Covered
0.47428	0.35051	0.00000	0.8	0.2	2.4	0.47500	Covered
0.52387	0.40070	0.00000	0.8	0.2	2.6	0.52500	Covered
0.57359	0.45078	0.00000	0.8	0.2	2.8	0.57500	Covered
0.62362	0.50069	0.00000	0.8	0.2	3	0.62500	Covered
0.12498	0.00293	0.00000	0.8	0.4	1	0.12500	Covered
0.17498	0.05011	0.00000	0.8	0.4	1.2	0.17500	Covered
0.22496	0.10011	0.00000	0.8	0.4	1.4	0.22500	Covered
0.27488	0.15019	0.00000	0.8	0.4	1.6	0.27500	Covered
0.32489	0.20014	0.00000	0.8	0.4	1.8	0.32500	Covered
0.37487	0.25013	0.00000	0.8	0.4	2	0.37500	Covered
0.42466	0.30029	0.00000	0.8	0.4	2.2	0.42500	Covered
0.47421	0.35056	0.00000	0.8	0.4	2.4	0.47500	Covered
0.52500	0.40000	0.00000	0.8	0.4	2.6	0.52500	Covered
0.57398	0.45057	0.00000	0.8	0.4	2.8	0.57500	Covered
0.62305	0.50097	0.00000	0.8	0.4	3	0.62500	Covered
0.12497	0.00387	0.00000	0.8	0.6	1	0.12500	Covered
0.17500	0.05001	0.00000	0.8	0.6	1.2	0.17500	Covered
0.22499	0.10003	0.00000	0.8	0.6	1.4	0.22500	Covered
0.27492	0.15014	0.00000	0.8	0.6	1.6	0.27500	Covered
0.32494	0.20007	0.00000	0.8	0.6	1.8	0.32500	Covered
0.37477	0.25023	0.00000	0.8	0.6	2	0.37500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.42459	0.30034	0.00000	0.8	0.6	2.2	0.42500	Covered
0.47436	0.35046	0.00000	0.8	0.6	2.4	0.47500	Covered
0.52435	0.40041	0.00000	0.8	0.6	2.6	0.52500	Covered
0.57411	0.45049	0.00000	0.8	0.6	2.8	0.57500	Covered
0.62375	0.50063	0.00000	0.8	0.6	3	0.62500	Covered
0.12500	0.00112	0.00000	0.8	0.8	1	0.12500	Covered
0.17499	0.05004	0.00000	0.8	0.8	1.2	0.17500	Covered
0.22498	0.10004	0.00000	0.8	0.8	1.4	0.22500	Covered
0.27496	0.15006	0.00000	0.8	0.8	1.6	0.27500	Covered
0.32490	0.20012	0.00000	0.8	0.8	1.8	0.32500	Covered
0.37491	0.25009	0.00000	0.8	0.8	2	0.37500	Covered
0.42472	0.30024	0.00000	0.8	0.8	2.2	0.42500	Covered
0.47455	0.35032	0.00000	0.8	0.8	2.4	0.47500	Covered
0.52449	0.40032	0.00000	0.8	0.8	2.6	0.52500	Covered
0.57379	0.45067	0.00000	0.8	0.8	2.8	0.57500	Covered
0.62397	0.50052	0.00000	0.8	0.8	3	0.62500	Covered
0.12500	0.00048	0.00000	0.8	1	1	0.12500	Covered
0.17498	0.05001	0.00006	0.8	1	1.2	0.17500	Covered
0.22498	0.10001	0.00004	0.8	1	1.4	0.22500	Covered
0.27498	0.15002	0.00002	0.8	1	1.6	0.27500	Covered
0.32497	0.20001	0.00003	0.8	1	1.8	0.32500	Covered
0.37496	0.25004	0.00000	0.8	1	2	0.37500	Covered
0.42487	0.30011	0.00000	0.8	1	2.2	0.42500	Covered
0.47481	0.35014	0.00000	0.8	1	2.4	0.47500	Covered
0.52464	0.40023	0.00000	0.8	1	2.6	0.52500	Covered
0.57409	0.45050	0.00000	0.8	1	2.8	0.57500	Covered
0.62405	0.50047	0.00000	0.8	1	3	0.62500	Covered
0.12500	0.00082	0.00001	1	0	1	0.12500	Covered
0.17493	0.05034	0.00000	1	0	1.2	0.17500	Covered
0.22485	0.10037	0.00000	1	0	1.4	0.22500	Covered
0.27476	0.15040	0.00000	1	0	1.6	0.27500	Covered
0.32439	0.20076	0.00000	1	0	1.8	0.32500	Covered
0.37427	0.25073	0.00000	1	0	2	0.37500	Covered
0.42500	0.30000	0.00000	1	0	2.2	0.42500	Covered
0.47427	0.35052	0.00000	1	0	2.4	0.47500	Covered
0.52416	0.40052	0.00000	1	0	2.6	0.52500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.57376	0.45069	0.00000	1	0	2.8	0.57500	Covered
0.61932	0.50283	0.00000	1	0	3	0.62500	Covered
0.12500	0.00007	0.00000	1	0.2	1	0.12500	Covered
0.17500	0.05001	0.00000	1	0.2	1.2	0.17500	Covered
0.22488	0.10030	0.00000	1	0.2	1.4	0.22500	Covered
0.27484	0.15026	0.00000	1	0.2	1.6	0.27500	Covered
0.32500	0.20000	0.00000	1	0.2	1.8	0.32500	Covered
0.37500	0.25000	0.00000	1	0.2	2	0.37500	Covered
0.42500	0.30000	0.00000	1	0.2	2.2	0.42500	Covered
0.47384	0.35083	0.00000	1	0.2	2.4	0.47500	Covered
0.52361	0.40087	0.00000	1	0.2	2.6	0.52500	Covered
0.57360	0.45078	0.00000	1	0.2	2.8	0.57500	Covered
0.62354	0.50073	0.00000	1	0.2	3	0.62500	Covered
0.12497	0.00405	0.00000	1	0.4	1	0.12500	Covered
0.17576	0.04602	0.00002	1	0.4	1.2	0.17500	Uncovered
0.22491	0.10023	0.00000	1	0.4	1.4	0.22500	Covered
0.27491	0.15016	0.00000	1	0.4	1.6	0.27500	Covered
0.32500	0.20000	0.00000	1	0.4	1.8	0.32500	Covered
0.37455	0.25045	0.00000	1	0.4	2	0.37500	Covered
0.42431	0.30058	0.00000	1	0.4	2.2	0.42500	Covered
0.47440	0.35043	0.00000	1	0.4	2.4	0.47500	Covered
0.52417	0.40052	0.00000	1	0.4	2.6	0.52500	Covered
0.57327	0.45096	0.00000	1	0.4	2.8	0.57500	Covered
0.62359	0.50071	0.00000	1	0.4	3	0.62500	Covered
0.12499	0.00272	0.00000	1	0.6	1	0.12500	Covered
0.17499	0.05006	0.00000	1	0.6	1.2	0.17500	Covered
0.22496	0.10011	0.00000	1	0.6	1.4	0.22500	Covered
0.27485	0.15025	0.00000	1	0.6	1.6	0.27500	Covered
0.32491	0.20011	0.00000	1	0.6	1.8	0.32500	Covered
0.37477	0.25023	0.00000	1	0.6	2	0.37500	Covered
0.42460	0.30033	0.00000	1	0.6	2.2	0.42500	Covered
0.47388	0.35080	0.00000	1	0.6	2.4	0.47500	Covered
0.52416	0.40053	0.00000	1	0.6	2.6	0.52500	Covered
0.57413	0.45048	0.00000	1	0.6	2.8	0.57500	Covered
0.62390	0.50055	0.00000	1	0.6	3	0.62500	Covered
0.12499	0.00217	0.00000	1	0.8	1	0.12500	Covered

Continued on next page

Table B.3 – continued from previous page

W_u	P_u	d_u	α	β	$\bar{\theta}$	W_c	$W_c \geq W_u$
0.17500	0.05001	0.00000	1	0.8	1.2	0.17500	Covered
0.22500	0.10001	0.00000	1	0.8	1.4	0.22500	Covered
0.27496	0.15007	0.00000	1	0.8	1.6	0.27500	Covered
0.32493	0.20008	0.00000	1	0.8	1.8	0.32500	Covered
0.37492	0.25009	0.00000	1	0.8	2	0.37500	Covered
0.42465	0.30029	0.00000	1	0.8	2.2	0.42500	Covered
0.47438	0.35044	0.00000	1	0.8	2.4	0.47500	Covered
0.52424	0.40047	0.00000	1	0.8	2.6	0.52500	Covered
0.57430	0.45039	0.00000	1	0.8	2.8	0.57500	Covered
0.62347	0.50077	0.00000	1	0.8	3	0.62500	Covered
0.12500	0.00000	0.00000	1	1	1	0.12500	Covered
0.17493	0.05033	0.00000	1	1	1.2	0.17500	Covered
0.22494	0.10016	0.00000	1	1	1.4	0.22500	Covered
0.27498	0.15000	0.00002	1	1	1.6	0.27500	Covered
0.32497	0.20004	0.00000	1	1	1.8	0.32500	Covered
0.37495	0.25005	0.00000	1	1	2	0.37500	Covered
0.42487	0.30011	0.00000	1	1	2.2	0.42500	Covered
0.47481	0.35013	0.00000	1	1	2.4	0.47500	Covered
0.52466	0.40022	0.00000	1	1	2.6	0.52500	Covered
0.57404	0.45053	0.00000	1	1	2.8	0.57500	Covered
0.62420	0.50040	0.00000	1	1	3	0.62500	Covered