Taxation and Privacy Protection on Internet Platforms *

Francis Bloch, Gabrielle Demange

Taxation and Privacy Protection on Internet Platforms[*]

Francis Bloch[†]   Gabrielle Demange[‡]

October 11, 2016

Abstract

This paper studies data collection by a monopolistic internet platform We show that the optimal strategy of the platform is either to cover the market or to choose the highest data exploitation level, excluding users with high privacy costs from the platform. For likely parameter values, the platform chooses an excessive level of data exploitation from the point of view of users. We study how different tax instruments can be used to reduce the level of data collection and analyze the effect of an opting-out option, letting users access the platform with no data collection. We show that a differentiated tax, taxing access revenues and data revenues at different rates is the most effective instrument and that the introduction of an opting-out option may harm users as it induces the platform to raise the level of data exploitation.

JEL classification numbers: H23, L86, L50

Keywords: Digital services, Privacy protection, Taxation, Opt-out and opt-in options

[†]Université Paris 1 and Paris School of Economics, 106-112 Boulevard de l'Hopital, 75647 Paris CEDEX 13, France. Email francis.bloch@univ-paris1.fr.

[‡]Paris School of Economics-EHESS, 48 Boulevard Jourdan, 75014 Paris, France . Email demange@pse.ens.fr.

1 Introduction

The precipitous decline in the cost of data collection and storage linked to the development of information technologies has transformed business models in advertising and commerce. While records on customers and sales histories have always existed, the digital economy now enables firms to exploit data at a much larger scale, opening up new opportunities for profit as well as new concerns about privacy and exploitation of personal data. Large sales platforms can now use detailed records of past sales histories to target users and engage in discriminatory dynamic pricing. Other platforms, like search engines or online social networks, use data on immediate search to auction off advertising spaces to clients, or sell search histories to intermediaries who accumulate data to better target users with ads. Business models of all giant internet platforms rely at different degrees on the collection and exploitation of personal data. The use of personal data is clearly one of the main specificities of the digital sector in modern industrial economies.

The development of "big data" and its potential exploitation raises two separate questions. First, data are a valuable input for internet platforms, but users voluntarily upload their data without any payment. One can argue that internet platforms are engaged in a barter agreement, where platforms deliver a valuable service (targeted proposals for products, targeted ads, outcomes of search, access to friends) in exchange for the uploading of data. But absent any price and financial transaction, it is difficult to assess whether this barter is "fair" and if users receive a fair share of the surplus. The immense profits of (some not all) internet platforms suggests that it may not be the case and that platforms benefit from a "free" input which is not paid at its true value. Furthermore, in the absence of financial transactions, governments cannot properly tax the benefit of personal data, creating a distortion with respect to other sectors, clouding the territoriality principle for the taxation of profits, and leading to extremely low levels of taxation of internet platforms through a clever use of transfer prices and the absence of records of financial transaction in countries where users reside. Second, users are rightly afraid that the collection of personal data infringes on their privacy. In addition, the resale of data to unknown intermediaries through opaque arrangements results in a loss of control on the dissemination of personal data to third parties. The exploitation of data, while it provides a valuable service to users by improving targeting, also necessarily involves a cost in privacy loss.

Even though the two problems of the absence of fair payment of data and privacy loss

seem unrelated at first glance, they are in fact closely connected. In this paper, we study how regulatory instruments, and in particular various forms of taxation, can be used to solve both problems at once.

We construct a model where users are differentiated only along their privacy cost. Their (identical) benefit of accessing the platform is made of two components, one derived from the access to the platform (say from the use of the search engine or the purchase of a good) and one which depends in an increasing way on the collection of data. Data collection is represented by a scalar, which might be interpreted as the duration of time during which personal histories are stored by the platform or the extent to which geo-localization is used. The collection of data enables the platform to better target offers to users and users to products or advertisers, resulting in an increase in the benefit to users as well as the value of users to advertisers or to the firm. Initially, we suppose, following current usage in the digital industry, that platforms do not charge users for their service and that their entire revenue comes from the other side of the market (advertisers or sales of future goods). A monopolistic platform chooses (and commits to) a degree of data exploitation balancing two effects: on the one hand, an increase in data collection increases revenues by increasing the value of the user to advertisers or for targeted pricing, on the other hand, an increase in data collection may deter users with high privacy cost to access the platform.

In this model, we first compute the optimal level of data exploitation chosen by the platform and the users. We show that the platform either chooses to cover the market making sure that all users access the platform, or chooses the maximal level of data exploitation, thereby excluding some users from the platform. The welfare of users is also maximized at one of two values: either at a low value in which all users access the platform (hence the value is lower than level of market coverage) or at the maximal degree of data exploitation. The optimal levels of data collection from the point of view of the platform and users are characterized as a function of the particular shape of the revenue and benefit functions. In the likely situation where the access benefit is large relative to the benefit of data collection or when the marginal benefit of data collection decreases sharply with the amount of data, the optimal level of data collection of the platform is excessive from the point of view of users. Hence the objective of the regulator is to select instruments to reduce the level of data collection and exploitation of the platform.

3

We then study how different forms of taxation affect data collection. We first observe that a tax on profits (or equivalently a tax on revenues because variable costs are negligible) does not affect the choice of the platform. A tax paid by the platform per user does not affect the marginal benefit of data exploitation, but it reduces the profit made on the marginal user accessing the platform, thereby reducing the cost of data collection. Hence, a tax per user (or per flow of data as users do not choose the level of data they upload), results in an increase in data exploitation. A specific tax paid by users (like a tax on internet service providers) reduces the level of data exploitation when the market is covered, but also increases the region of parameters for which exclusion occurs. Hence, it produces ambiguous effects on the welfare of users. The only tax which allows to correct for excessive data collection is a tax on revenues which treats differentially platform's revenues accruing from one-time use (like auction revenues based on current keywords) and revenues linked to data collection (like resale of data to intermediaries). If fiscal authorities charge a higher tax level on resale of data than on auction revenues, taxes deter the platform from exploiting the data, playing the classical role of a Pigovian tax correcting for externalities.

We then explore the effect of the introduction of an option for the users to access the platform with no data collection. This allows the platform to collect access revenue from those users who choose to opt-out, segmenting the market into two groups with different revenue levels. Users will now all access the platform – some with data collection and others without. We show that this changes the level of data collection chosen by the platform, resulting in a decrease in the maximal level of data exploitation for which the market is covered, but in an increase in the region of parameters for which the platform chooses the maximal degree of data exploitation. We show that for the most likely values of the parameters, the platform's profit decreases with the introduction of the opt-out option, as this creates competition with the regular service with data collection and hence reduces the platform's ability to generate data revenues. More surprisingly, we show that users may also be harmed, because the introduction of the opt-out option may lead the platform to switch from a r/'egime of market coverage to the maximal level of data exploitation, hurting users with low privacy costs who choose to access the platform with data collection.

Our analysis of the effect of taxation and regulation on data collection is related to two

strands of the literature.

First, it is related to the literature on the economics of media, which considers a media (television, newspaper) as a platform in a two-sided market connecting readers with advertisers and collecting revenues from the advertising market. (See Gabsewicz, Laussel and Sonnac (2001) and (2004) for early contributions to the literature and the survey by Anderson and Gabszewicz (2006)). Advertisement in these models play the same role as data collection in ours. As in our model, users are assumed to suffer a linear cost from advertisements. Though there are two main differences between our analysis and this literature. One stems from the questions raised, the other on the modelling assumptions. The recent literature on media as two-sided platforms mostly focuses on program differentiation and competition, while we are interested in regulatory régimes and taxation of a monopolistic platform to improve privacy protection. For instance Peitz and Valletti (2008), Choi (2006), Crampes, Haritchabalet and Jullien (2009)) allow for subscription prices charged to viewers, and compare regimes of "free-to-air" with "pay-for-view" televisions. Reisinger, Ressner and Schmidtke (2009) and Reisinger (2012) pay close attention to competition among advertisers on different platforms. Casadesus-Masanell and Hervas-Drane (2015) analyze a model where consumers choose the amount of data they upload and platforms choose a two-pronged policy to extract access and data revenues. They show that platforms typically specialize in access or data collection revenues. The papers that consider a monopolistic platform rely on different modeling assumptions, which precludes a direct comparison between our results and theirs (see e.g. Anderson and Gans (2011) and Johnson (2013) who examine the reaction of a platform to ad-blocking technologies). For example Anderson and Coate (2005) analyze the behavior of a monopolistic platform when the market is not covered, in their model, users are differentiated by their intrinsic benefit from the good and not their aversion to advertising, resulting in very different demand functions and different conclusions from here. In particular, they find that platforms typically choose too little advertising whereas we observe that the level of data collected by the platform is likely to be excessive. Tag (2009) introduces the possibility for consumers to pay to pay to remove advertisements. Almost all models assume that users suffer a linearly increasing utility loss for ads, whereas we assume that, in addition to a linearly decreasing loss, they obtain a concave utility gain due to improved service. Crampes, Haritchabalet and Jullien (2009) however allow users to have a positive value for low levels of

ads, which is similar but not equivalent to our assumption. De Cornière and de Nijs (2016) also discuss privacy in a monopolistic two-sided platform, but with a different angle. They study how the platform's choice between privacy and disclosure affect competition among advertisers.

Second, our paper is more distantly related to the literature on taxation on two-sided markets (see Kind et al. (2008), (2010a), (2010b), Kind et al. (2013) and Kotsogiannis and Serfes (2010)), but again the focus of the analyses are different, as we focus on the effect of taxes on privacy protection rather than revenues and distortion. Finally, we note that our paper is closely related to the two contributions by Crémer (2015) and Bourreau, Caillaud and De Nijs (2015) as all three papers consider the effect of specific taxation on the behavior of internet platforms, even though the three papers rely on different models of internet platforms and focus on different choice variables.

The rest of the paper is organized as follows. We introduce our model in the next Section. Section 3 is devoted to the baseline model. Section 4 analyzes the effect of different tax instruments. Section 5 considers the introduction of an option to access the platform with no data collection. Section 6 concludes. All proofs not given in the texts are collected in Section 7.

2 The Model

2.1 Platform and users

The platform We consider an internet platform which provides services to users. The platform collects revenues either directly (as in the case of e-commerce) or from third parties like advertisers in the case of search engines or digital social networks. The platform also collects data from users, recording their history of activity on the platform. We distinguish between two sources of revenues: some of the revenues, like sales revenues from e-commerce or instantaneous search-base advertising revenues are collected immediately, whether the platform records data on the customer or not. Other revenues, like the sale of personal data to aggregators or the expected revenue from future targeted advertising, only accrue if personal data are collected and stored. We suppose that the platform commits to the degree of exploitation of personal data, denoted $x \in [0, 1]$. The degree x can be interpreted along different dimensions. It can represent the duration of time during which personal histories are stored by the platform, the fraction of

personal data which are sold by the platform to third parties or kept for direct exploitation, or any specified limitation on the use of personal data. We denote by v_0 the fixed value generated by a user independently of data collection (the access revenue) and by $v(x)$ the value generated by data exploitation (data revenue). The function $v(\cdot)$ is supposed to be iso-elastic so that the total value of a user to the platform is

$$v_0 + v(x) \text{ where } v(x) = bx^{\beta} \text{ for some } b > 0, 0 < \beta < 1.$$

Finally, we assume the marginal cost of providing digital services is negligible.

Users Users get a benefit from using the platform, which is also decomposed into a fixed positive component u_0 (the access benefit), and a component which depends on the collection of data x (the data benefit). The fixed component u_0 corresponds to the benefit from accessing the platform, i.e. buying a good, obtaining results from a query on a search engine, contacting friends on social networks. Data collection allows the platform to better match the user with a product and hence results in an improvement in the service. We let $u(x)$ denote the benefit from the improvement in match quality, where $u(\cdot)$ is supposed to be iso-elastic. In addition, we suppose that users are aware that they may be harmed by the collection of storage and personal data, and suffer a privacy cost which we assume proportional to the degree of data exploitation. While users are identical in all other characteristics (in particular their value to advertisers and the benefit they obtain from the platform), we suppose that they are heterogeneous in their sensitivity to privacy loss. We let θ denote the characteristic of the user with privacy cost θx, called a θ-user. We assume the distribution of privacy costs in the population of users, θ, to be uniform over $[0, 1]$. The utility of a θ-user can be written as

$$u_0 + u(x) - \theta x \text{ where } u(x) = ax^{\alpha} \text{ for some } a > 0, 0 < \alpha < 1.$$

We normalize the payoff of a user who does not access the platform to zero.

In the baseline scenario, users access the platform letting the platform use their data. Given the positive level of data exploitation x, a θ-user chooses to access the platform if and only if

$$u_0 + u(x) - \theta x \geq 0.$$

For x small enough,[1] $u_0 + u(x) - x \geq 0$ so that all users access the platform. Otherwise, $u_0 + u(x) - x < 0$, and the characteristic of the marginal user accessing the platform is given by the solution to the equation: $u_0 + u(x) - \theta x = 0$. So letting $T(x)$ defined by

$$T(x) = \max(\frac{u_0 + u(x)}{x}, 1) \qquad (1)$$

all users with characteristic in $[0, T(x)]$ access the platform and the demand is given by $T(x)$.

The assumptions of iso-elasticity on the data revenue and user's benefit functions can be relaxed as discussed at the end of Sections 3 and 4.

Our model can be compared to models of interaction on media markets when users are averse to advertisement. In these models, x is interpreted as the level of advertising, $v(x)$ the revenues from advertising and θx the cost of ads to users. Models of advertising in media markets typically assume that $v_0 = 0$ (the media does not make any direct revenue from users) and that $u(x) = 0$ (users have no positive value for advertising).[2] A more important difference stems from the dimension of heterogeneity of users. We assume that users have the same value for the service but different privacy costs, whereas models of media markets suppose that users have the same cost of advertising but differ in their value of the service or in their ideal point in models of horizontal differentiation across media. This difference between our model and other models of media with advertising avoidance results in non-trivial consequences. When users differ in their value for the service or ideal points, demand is linearly decreasing in x ; when users differ in their advertising or privacy cost, demand is no longer linear and in fact includes a part $\frac{u_0}{x}$ which is decreasing and convex in x. Hence profit is typically concave in models with heterogeneity in values, but not in models with heterogeneity in sensitivity to advertising or privacy. This leads to a major difference in the characterization of optimal levels of advertising or data exploitation.

2.2 Regulatory instruments

The objective of the paper is to assess how different regulatory instruments affect the level of data collection chosen by the platform and the welfare of users. We distinguish between tax

---

[1]The function $\frac{u_0 + u(x)}{x}$ is decreasing in x.
[2]One exception is Crampes et al. (2009) who allow users to have a positive value for low levels of advertising as we do.

instruments and a regulation imposing an opt-out option. We investigate the incidence of four different tax instruments:

- A tax τ levied on the revenue of the platform

- A tax t_P levied on the platform per user

- A differentiated revenues tax system, with two different tax rates τ_1 and τ_2 applied to the revenues v_0 and $v(x)$

- A tax t_U levied on each user for accessing the platform.

As an alternative regulation, we consider a policy mandating that the platform offer a choice to the user between access with data collection and an opt-out option without data collection. We focus on the case where the opt-out option is offered at no charge to the user.

3   The baseline model

In the baseline model, there is no public intervention and the platform freely chooses the degree of data exploitation x. Our aim in this Section is to compare the optimal level of data exploitation chosen by the platform and the users. The profit of the platform is given by

$$\Pi(x) = [v_0 + v(x)]T(x). \tag{2}$$

The user's welfare is defined as the sum of the utility levels of all users. As the benefit of users who do not access the platform is normalized to zero, the user's welfare with data collection x is given by

$$W(x) = \int_0^{T(x)} [u_0 + u(x) - \theta x] d\theta. \tag{3}$$

For x low enough, all users access the platform (for $u_0 > 0$). As x increases, users with high privacy cost stop accessing the platform, and the demand becomes decreasing in x. We define the *coverage level* to be the maximal value of data exploitation for which all users access the platform.

9

The coverage level The coverage level will be used to characterize the optimal behavior of the platform and discuss the effect of regulatory instruments. It is convenient to emphasize the dependance of the coverage level with respect to the access benefit u_0. As a preliminary Lemma, we derive some comparative statics properties of the coverage level ξ.

Property 1 *Let $\xi(u_0)$ be the coverage level: $T(x) = 1$ if and only if $x \leq \xi(u_0)$.*

If $u_0 + a > 1$, the coverage level $\xi(u_0)$ is 1: all users access the platform at the maximal degree of data exploitation $x = 1$.

If $u_0 + a \leq 1$, the coverage level $\xi(u_0)$ is the unique solution in $(0, 1)$ to the equation:

$$u_0 + ax^\alpha - x = 0. \tag{4}$$

ξ is increasing in u_0; when u_0 tends to 0, $\xi(u_0)$ tends to $\xi(0) = a^{\frac{1}{1-\alpha}}$; $\xi(1-a) = 1$. ξ is increasing in a and decreasing in α.

Not surprisingly, an increase in the access benefit and a uniform increase in the data benefit increase the utility of users, and raises the coverage level. The effect of a change in α is less obvious. An increase in α makes the benefit less concave in data and hence reduces the marginal benefit of data for low values of x, lowering the coverage level.

In the sequel of the paper, we consider the interesting case where the coverage level is interior and suppose that $u_0 + a < 1$.

3.1 The platform's choice

To compute the optimal level of data collection of the platform, we observe that the expression of the platform's profit differs in the régime where the market is covered, when $x \leq \xi(u_0)$ and where some users do not access the platform, when $x > \xi(u_0)$:

$$
\begin{aligned}
\Pi(x) &= v_0 + v(x) \text{ if } x \leq \xi(u_0), \\
&= [v_0 + v(x)]\frac{u_0 + u(x)}{x} \text{ if } x \geq \xi(u_0).
\end{aligned}
$$

When the level of data exploitation is below the coverage level, the profit is strictly increasing in x because the demand stays at the maximum level. When the level of data exploitation is above

the coverage level, the profit might no longer be increasing because the demand $T(x) = \frac{u_0 + u(x)}{x}$ is decreasing in x. With our iso-elastic specification, the elasticity of the data benefit function, $u(\cdot)$, is constant and smaller than 1. The marginal data benefit thus $u'(x)$ decreases less rapidly than the average benefit $\frac{u(x)}{x}$, so that the demand $T(x)$ decreases slowly with an increase in x. Similarly, the marginal value of data exploitation $v'(x)$ decreases less rapidly than the average value $\frac{v(x)}{x}$. This implies a weak form of convexity for the profit, namely that the elasticity of profit with respect to data exploitation, $\frac{x\Pi'(x)}{\Pi(x)}$ is increasing in x. As a result, the platform will never find it optimal to choose an intermediate value and the optimal choice is either equal to $\xi(u_0)$ or 1.

There are thus two possible régimes depending on the values of the parameters: a régime of market coverage where the platform chooses $x = \xi(u_0)$ and serves all users and a régime of exclusion where the platform selects $x = 1$ and serves users with a privacy cost smaller than $u_0 + a$. The values of the parameters determining each régime are given in next Proposition.

Proposition 1 *The platform optimally chooses a degree of data exploitation equal either to the coverage level $\xi(u_0)$ or to the maximal level 1. It chooses $\xi(u_0)$ if and only if $\frac{v_0}{b} \geq \nu(u_0) \equiv \frac{u_0 + a - \xi(u_0)^\beta}{1 - u_0 - a}$. The threshold $\nu(u_0)$ is increasing in α and β and decreasing in u_0.*

Proposition 1 establishes the existence of a threshold $\nu(u_0)$ such that the platform chooses market coverage when the market when the fraction $\frac{v_0}{b}$ is larger than $\nu(u_0)$ and exclusion otherwise. This result is easily understood. The fraction $\frac{v_0}{b}$ measures the relative importance of the access revenue over the data revenue due. When the access revenue is large relative to the data revenue, the platform has an incentive to serve all users and hence will choose x low enough so that it covers the market. If the value of data exploitation is large relative to the value of access, the opposing intuition holds and the platform optimally chooses to restrict access and set a high degree of data exploitation. When the access benefit u_0 increases, the coverage level $\xi(u_0)$ goes up, making market coverage more attractive to the platform. By contrast, an increase in α makes the data benefit less concave and reduces the coverage level, making exclusion more attractive. When β increases, the data revenue becomes less concave, increasing the effect of a change in x on the platform's profit, and thereby making exclusion more attractive.

These comparative statics effects of α and β can be illustrated in the simple case where $u_0 = 0$. We compute $\nu(0) = \frac{a(1 - a^{\alpha + \beta - 1})}{1 - a}$. As $a < 1$, the sign of $\nu(0)$ is equal to the opposite

of the sign of $\alpha + \beta - 1$, so that $\nu(0) < 0$ if $\alpha + \beta < 1$. As the threshold is decreasing in u_0, we obtain the following corollary, characterizing conditions under which market coverage and exclusion are chosen by the platform:

Corollary 1 *For $\alpha + \beta \leq 1$, the platform always chooses market coverage. For $\alpha + \beta > 1$, there exists a threshold of the access benefit, $\overline{u_0}$ such that the platform chooses market coverage for $u_0 \geq \overline{u_0}$ and market exclusion for $u_0 < \overline{u_0}$.*

3.2 Users' welfare

The value of users' welfare also assumes two different expressions, depending on whether market is covered or not:

$$
\begin{aligned}
W(x) &= u_0 + u(x) - \frac{x}{2} \text{ if } x \leq \xi(u_0) \\
&= \frac{1}{2x}[u_0 + u(x)]^2 \text{ if } x \geq \xi(u_0).
\end{aligned}
$$

When the level of data exploitation is below the coverage level, all users access the platform, and welfare is strictly concave in x. When the level of data exploitation is above the coverage level, the expression for welfare captures the fact that some users are excluded from the platform, and welfare is no longer necessarily concave in the level of data exploitation x. We note that, at the boundary value $\xi(u_0)$, the utility of a marginal user is identical in the two régimes so that the welfare function $W(x)$ is differentiable. Thus, contrary to the platform's profit, the welfare function does not present a kink at the coverage level and the optimal level of data collection will generically not be equal to the boundary value $\xi(u_0)$.

Let $\widehat{x}$ denote the value which maximizes the welfare of the average user, $u_0 + u(x) - \frac{x}{2}$. With our specification, $\widehat{x} = (2a\alpha)^{\frac{1}{1-\alpha}}$. When $\widehat{x}$ is above the coverage level, welfare increases with x up to the coverage level $\xi(u_0)$: covering the market is not optimal. When the level of data exploitation is above the coverage level, the market is not covered, and the demand decreases slowly with x so that the welfare function displays the same weak form of convexity as the profit: the elasticity of welfare with respect to the level of data exploitation, $\frac{xW'(x)}{W(x)}$ is increasing in x, which shows that the optimal choice cannot be an intermediate value between $\xi(u_0)$ and 1.

We conclude that the optimal choice of users will either be equal to $\widehat{x}$ (only if this value is

lower than $\xi(u_0)$) or to 1.

Proposition 2 *The level of data collection maximizing users' welfare is either $\widehat{x}$ or 1. If $\alpha \leq \frac{1}{2}$, users prefer market coverage at $\widehat{x}$. If $\alpha \geq \frac{1}{2}$, there exists a threshold value of the access benefit, $\widehat{u_0}$ such that users prefer market coverage at $\widehat{x}$ if $u_0 \geq \widehat{u_0}$ and prefer exclusion if $u_0 < \widehat{u_0}$.*

Proposition 2 characterizes the optimal level of data collection from the users point of view. We first observe that when u_0 increases, the difference $W(\widehat{x}) - W(1)$ increases: when the access benefit is higher, exclusion becomes less attractive and users prefer market coverage to exclusion. We now focus attention to the case $u_0 = 0$, and observe that (i) $\widehat{x} \leq \xi(u_0)$ if and only if $\alpha \leq \frac{1}{2}$ and (ii) for any $x > \xi(u_0)$, $W'(x) \leq 0$ if and only if $\alpha \leq \frac{1}{2}$. The value $\alpha = \frac{1}{2}$ plays a special role in the computations: when the data revenue function is a square root function, the welfare $W(x)$ is increasing up to $\xi(u_0) = \widehat{x} = a^2$, and equal to $\frac{a^2}{2}$ afterwards. When the data benefit function is more concave than the square root function ($\alpha < \frac{1}{2}$), the maximum is obtained for a value lower than $\xi(u_0)$, and the welfare function is decreasing after $\xi(u_0)$. By contrast, when the data benefit function is less concave than the square root function, ($\alpha > \frac{1}{2}$), the maximum is achieved at a value $\widehat{x} > \xi(u_0)$ and the welfare function is increasing after $\xi(u_0)$. We conclude that, when $u_0 = 0$, if $\alpha \leq \frac{1}{2}$, users prefer market coverage to exclusion, but if $\alpha \geq \frac{1}{2}$, users prefer exclusion to market coverage. As an increase in u_0 makes market coverage more attractive, we know that even when $\alpha \geq \frac{1}{2}$, market coverage will be preferred to exclusion when the access benefit is sufficiently large.

Propositions 1 and 2 show that the optimal level of data collections of the platform and of users may differ and that their ranking depends on the parameters. These Propositions cannot rule out situations where the platform chooses market coverage and the users prefer exclusion (this may happen for low values of u_0 when $\alpha > \frac{1}{2}$ and $\alpha + \beta < 1$). However, we believe that such situations are unlikely to arise in practice as it requires access benefits to be small with respect to data benefit, and the data benefit function to be more elastic than the data revenue function, implying that advertisers are less sensitive to incremental increases in data than users. If $\alpha \leq \frac{1}{2}$ or $u_0 \geq \widehat{u_0}$, users always favor market coverage at a level of data collection $\widehat{x} \leq \xi(u_0)$ so that *users always favor a level of data exploitation which is lower than the platform's.* We conjecture that this is the case which is empirically relevant, as it corresponds to an access benefit which

is high with respect to the data benefit, and a data benefit function which achieves most gains for low levels of data exploitation. We state this result in the following corollary.

Corollary 2 *If $\alpha \leq \frac{1}{2}$ or $\alpha \geq \frac{1}{2}$ and $u_0 \geq \widehat{u_0}$ users favor a level of data exploitation $\widehat{x}$ which is lower than $\xi(u_0)$, hence surely lower than the platform's choice. In that case welfare decreases when x increases from $\widehat{x}$ to $\xi(u_0)$.*

Robustness   We now argue that Propositions 1 and 2 can be extended to data benefit and value functions which are more general than the iso-elastic functions. Our observation that the optimal level of data exploitation of the platform and of users cannot be an interior point in the segment $[\xi(u_0), 1]$ relies on the fact that the elasticity of profit and welfare is non-decreasing in the region where $T(x) = \frac{u_0 + u(x)}{x} < 1$. A sufficient condition for the elasticity of profit to be non-decreasing is that both the elasticities of the $u(\cdot)$ and $v(\cdot)$ functions be non-decreasing, as

$$\frac{x\Pi'(x)}{\Pi(x)} = \frac{xv'(x)}{v_0 + v(x)} + \frac{xu'(x)}{u_0 + u(x)} - 1.$$

Similarly, a sufficient condition for the users' welfare to be non-decreasing is that the elasticity of the data benefit function $u(\cdot)$ be non-decreasing as

$$\frac{xW'(x)}{W(x)} = 2\frac{xu'(x)}{u_0 + u(x)} - 1.$$

These conditions are clearly satisfied for iso-elastic functions as $\frac{xv'(x)}{v(x)} = \beta$ and $\frac{xu'(x)}{u(x)} = \alpha$. We also note that the observation that the welfare function is differentiable at $\xi(u_0)$ does not depend on the specification of the data benefit function. Hence, the characterization of the optimal level of data collection of the platform as either the coverage level $\xi(u_0)$ or the maximal level 1 and of the optimal level of data collection of users as either the optimal level $\widehat{x}$ for the average consumer or 1 does not rely on the specification of the functions $u(\cdot)$ and $v(\cdot)$.

4   Taxation

In this Section, we analyze the effect of the imposition of a tax on the strategy of the platform. The imposition of different taxes can be interpreted as changes in the parameters of the model.

A tax τ on the revenues only scales down the profit and has no effect on the strategy of the platform. A tax per user paid by the platform reduces the access revenue from v_0 to $v_0 - \tau_P$. A differentiated revenues tax reduces the access revenue from v_0 to $v_0 - \tau_1$ and the data revenue from $v(x)$ to $v(x)(1 - \tau_2)$. A tax per user paid by users reduces the access benefit from u_0 to $u_0 - \tau_U$. The effect of taxes on the optimal choice of the platform can thus be computed by using the comparative statics results of the previous Section. They are summarized in the following Proposition.

Proposition 3 *The imposition of a tax has the following effects on the degree of data exploitation by the platform:*

- *An ad valorem tax τ on the profits or revenues of the platform has* no effect *on the degree of data exploitation*

- *A tax paid by the platform per user τ_P results in an* increase *in the degree of data exploitation*

- *In a differentiated revenues tax system, a tax τ_1 on the access revenues results in an* increase *in the degree of data exploitation, whereas a tax τ_2 on the revenues generated by data exploitation results in a* decrease *in the degree of data exploitation*

- *A tax paid by users for accessing the platform τ_U results in a decrease in the degree of data exploitation when the platform chooses market coverage, but may increase the probability that the platform chooses exclusion.*

Proposition 3 analyzes the incidence of the different taxes on the optimal choice of the platform. Clearly, a uniform tax on profits, or on revenues given that marginal cost are negligible, is neutral, and does not affect the platform's choice of data exploitation. A tax per user paid by the platform has no effect on the marginal data revenue of users accessing the platform, but it lowers the loss of not serving the marginal user who chooses not to access on the platform. Hence a tax per user induces the platform to *increase* its level of data exploitation. Similarly, in the context of differentiated taxation on access and data revenues, a tax τ_1 on the access revenue leaves the marginal benefit of data exploitation unchanged but reduces the loss of not serving the marginal user and hence raises data exploitation. By contrast, a tax τ_2 targeted at

data revenues leads to a reduction in the elasticity $\frac{xv'(x)}{v_0+v(x)}$ while leaving the effect on demand unchanged. It thus reduces the platform's incentive to collect data and results in a lower level of data exploitation. Finally, a tax supported by users lowers the level of data exploitation when the platform chooses market coverage. However, it also makes market coverage less attractive and hence increases the parameter configurations for which the platform chooses exclusion. However, recall that when the data revenue and benefit functions are highly inelastic, $\alpha+\beta<1$, the platform always chooses to cover the market. In that case, a tax on users unambiguously leads to a reduction in data exploitation.

User's welfare Proposition 3 enables us to study the effect of taxation on user's welfare through the change in the level of data exploitation in the likely situation identified in Corollary 2 where users prefer a level of data exploitation $\widehat{x}$, which is lower than $\xi(u_0)$.

Corollary 3 *If $\alpha \leq \frac{1}{2}$ or $\alpha \geq \frac{1}{2}$ and $u_0 \geq \widehat{u_0}$, any taxation which raises the level of data exploitation hurts users and any taxation which lowers the level of data exploitation, keeping it above[3] $\widehat{x}$, benefits users.*

This result does not compare the different tax instruments. Different taxation schemes may generate different fiscal revenues that could be incorporated in the measure of social welfare.

Robustness Finally we extend the tax incidence results of Proposition 3 to general utility functions with non-decreasing elasticity (for which the the same types of results obtain, see the discussion at the end of Section 3). Proposition 3 only depends on the comparative statics effects of changes in u_0, v_0 and b on the optimal level of data exploitation of the platform. Now note that $\Pi(\xi(u_0)) - \Pi(1) = v_0(1 - T(1)) + b(v(\xi(u_0)) - T(1)v(1))$ is increasing in v_0; it is decreasing in b if $x = 1$ is preferred by the platform since then surely $v(\xi(u_0)) - T(1)v(1))$ is negative or any functions $v(\cdot)$ (still assuming that $T(1) < 1$, i.e. not everyone access the platform if the maximum level is chosen). In addition, the coverage level $\xi(\cdot)$ is an increasing function, by the

---

[3]Welfare increases when x decreases down to $\widehat{x}$. When $\alpha \leq \frac{1}{2}$, then $\widehat{x} < \xi(0)$, so that the platform's choice stays always larger than $\widehat{x}$: users surely benefit from a decrease in the level of data exploitation. When $\alpha \geq \frac{1}{2}$ and $u_0 \geq \widehat{u_0}$ the decrease in x might be harmful if the platform chooses a level below $\widehat{x}$. But this can happen only if the tax is high enough to have a large effect.

concavity of u[4] These observations are sufficient to guarantee that Proposition 3 carries over to general utility functions, as the marginal effects of an increase in u_0 and v_0 on the behavior of the platform do not depend on the iso-elastic specification.

5 Opt-out option

In this section, we suppose that the platform is mandated to propose an opt-out option, where users are guaranteed that data will not be collected but suffer from the resulting loss in service quality. For example, users may refuse to be geo-localized or refuse to let cookies monitor their activity, but will in consequence obtain a lower quality of service on the platform. Users now freely choose between two options: one with no data collection and one with a single positive level x of data collection.[5] As each user prefers the opt-out option to the outside option ($u_0 > 0$) all users access the platform. We examine the optimal level of data collection with an opt-out option, compare the profit of the platform and the welfare of users in the baseline model and with the opt-out option.

A θ-user chooses the positive level of data collection x instead of the opt-out option if and only if

$$u(x) + u_0 - \theta x \geq u_0,$$

or $\theta < \frac{u(x)}{x}$. With the opt-out option, the coverage level is the solution to

$$ax^\alpha - x = 0,$$

or $\xi(0) = a^{\frac{1}{1-\alpha}}$, the *coverage level defined in the baseline model when $u_0 = 0$*. The profit of the platform is

$$
\begin{aligned}
\Pi(x) &= v_0 + v(x) \text{ if } x \leq \xi(0), \\
\Pi(x) &= v_0 + v(x)\frac{u(x)}{x} \text{ if } x \geq \xi(0).
\end{aligned}
$$

---

[4]($\xi(u_0)$ satisfies the equation. Thus surely $u(x) \leq x$ at $x = \xi(u_0)$, which implies $u'(x) < 1$. Applying the implicit function theorem thus yields that $\xi(\cdot)$ is an increasing function..

[5]For reasons of tractability and technical implementability, we do not consider more complex policies where the platform discriminates among users by proposing several positive degrees of data collection.

Notice that the platform now collects the access revenue v_0 on all users, and that the demand for data collection is independent of the access benefit u_0. The optimal level of data collection of the platform is thus easier to characterize than in the baseline model, as it corresponds to the level chosen by the platform when $u_0 = 0$. The profit $\Pi(x)$ is increasing in x for $x \leq \xi(0)$ and either increasing or decreasing in x for $x \geq \xi(0)$ depending on whether $\alpha + \beta$ exceeds 1. We immediately obtain the following characterization:

Proposition 4 *With an opt-out option, the platform optimally chooses a degree of data exploitation $x \in \{\xi(0), 1\}$. It chooses $\xi(0)$ if and only if $\alpha + \beta \leq 1$.*

Proposition 4 shows that, as in the baseline case, the platform either chooses to cover the market or sets the highest possible degree of data exploitation. It covers the market when the data revenue and benefit functions are highly concave, $\alpha + \beta \leq 1$ and chooses maximal data exploitation otherwise. The comparison between the level of data collection of the platform with and without the opt-out option produces ambiguous results. On the one hand the coverage level $\xi(0)$ is always lower than the coverage level in the baseline case, $\xi(u_0)$. The existence of the opt-out option forces the platform to reduce x in order to attract all users. On the other hand, the region of parameters for which the platform chooses maximal data exploitation may increase. The platform now collects the access revenue v_0 on all users and does not face the risk of non-participation as in the baseline case. When $\alpha + \beta > 1$ and u_0 is large, this effect leads the platform to choose maximal exploitation when the opt-out option is present, whereas it chooses to cover the market in the baseline case.

We next characterize the optimal level of data exploitation from the users point of view. An easy adaptation of Proposition 2 when $u_0 = 0$ yields the following result.

Proposition 5 *With an opt-out option, the welfare maximizing level of data collection is $\widehat{x}$ if $\alpha \leq \frac{1}{2}$ and 1 if $\alpha \geq \frac{1}{2}$.*

With an opt-out option, all users access the platform and obtain the access benefit u_0. Thus the optimal value of data collection is not affected by u_0, and is the same as the optimal level in the baseline case when $u_0 = 0$. Not surprisingly, exclusion is optimal more often than under a uniform policy. Comparing Propositions 2 and 5, we see that when $\alpha > \frac{1}{2}$ and u_0

18

is large, users favor the maximal level of data exploitation when the outside option is present whereas they prefer market coverage in the baseline mode. Hence the presence of the outside option unambiguously increases the optimal level of data exploitation of the users, because this benefits those who do not value privacy much while leaving the access benefit to others.

We next compare the profit of the platform in the baseline case and with the opt-out option.

Proposition 6 *The profit of the platform is higher in the baseline case than with the outside option whenever $\alpha + \beta < 1$. If $\alpha + \beta \geq 1$, there exists a threshold value $\tilde{u}_0 \in (0, 1-a)$ of the access benefit such that the profit of the platform is higher in the baseline case than with the outside option if and only if $u_0 > \tilde{u}_0$.*

Proposition 6 shows that the only situation where the platform prefers to introduce the opt-out option is the unlikely situation where the access benefit is low and the data revenue and benefit functions are elastic enough. In all other circumstances, the platform is hurt by the introduction of the opt-out policy. The intuition underlying Proposition 6 is easy to grasp. When $\alpha + \beta < 1$, the platform covers the market at level $\xi(0)$, which is lower than $\xi(u_0)$. Given that, profit is increasing in x under market coverage, this policy is always dominated by a market coverage policy at $\xi(u_0)$ which can be chosen by the platform in the baseline case. When $\alpha + \beta \geq 1$, we observe that the profit of the platform with the outside option is independent of u_0. By contrast, the profit of the platform is strictly increasing in u_0 in the baseline scenario. At $u_0 = 0$, the platform chooses exclusion in both régimes and the profit of the platform is strictly higher with the opt-out option as it collects access revenue v_0 on all users. On the other hand at $u_0 = 1 - a$, the profit of the platform is higher in the baseline model, as it attracts all users by setting the maximal level of data collection. Hence there exists a threshold value $\tilde{u}_0 \in (0, 1-a)$ under which profit is higher with the outside option and over which profit is higher in the baseline case.

We now ask whether users benefit from the switch from the introduction of the outside option. The answer to this question is not easy. On the one hand, if the baseline policy involves exclusion, excluded users benefit from the introduction of the opt-out option. On the other hand, the introduction of the opt-out option may lead to a deviation of the level of data collection away from the optimum, hurting users who access the platform. The following Proposition details situations under which the introduction of the outside option benefits users.

Proposition 7 *Suppose that $\alpha \leq \frac{1}{2}$. If $\alpha + \beta \leq 1$ or $\alpha + \beta \geq 1$ and $u_0 < \overline{u_0}$, users benefit from the introduction of the opt-out option. If $\alpha + \beta \geq 1$ and $u_0 > \overline{u_0}$, users are harmed by the introduction of the opt-out option if and only if $\xi(u_0) < a^2$.*

Proposition 7 focuses on the situation $\alpha \leq \frac{1}{2}$, where users' optimal level of data collection is lower than the platform's choice both in the baseline model and with the outside option. If the platform chooses market coverage in both régimes (i.e. when $\alpha + \beta < 1$), users prefer the situation with the opt-out option as the coverage level is lower than in the baseline model. Similarly, users prefer the situation with outside option when the platform chooses the maximal level of data exploitation under the two régimes (i.e. when $\alpha + \beta > 1$ and $u_0 > \overline{u_0}$) as excluded users obtain a positive benefit with the opt-out option. However, when the platform chooses to cover the market in the baseline model but to set the maximal level of data collection in the presence of the outside option (i.e. when $\alpha + \beta > 1$ and $u_0 > \overline{u_0}$), users with low privacy cost who choose data collection are harmed and users' welfare may be higher in the régime without outside option. Hence, for the most likely values of the parameters, the introduction of an opt-out option reduces the profit of the platform, but does not necessarily have a positive impact on users, as it may lead the platform to switch from market coverage to the maximal level of data exploitation.

6   Conclusion

This paper studies data collection by a monopolistic internet platform We show that the optimal strategy of the platform is either to cover the market or to choose the highest data exploitation level, excluding users with high privacy costs from the platform. For likely parameter values, the platform chooses an excessive level of data exploitation from the point of view of users. We study how different tax instruments can be used to reduce the level of data collection and show that user-based taxes lead to an increase in data collection. Taxation with different rates according to the source of revenues, with higher tax level for revenues generated by data exploitation can reduce data collection as can a tax paid by users on access. We also analyze the effect of an opting-out option, letting users access the platform with no data collection and show that the

platform is hurt by the introduction of the outside option and that, more surprisingly, users may also be harmed as the level of data exploitation increases.

Our analysis provides a first step in understanding how different regulatory instruments affect the optimal degree of data collection of an internet platform (or equivalently the optimal level of advertising of a media platform). Our analysis is based on a simple model of behavior of platforms and users and three directions seem particularly fruitful to follow in future research. First, we wish to study how another dimension of heterogeneity among users (for example demographic characteristics and the frequency of access) affects our results. Second, we would like to open the black box of the value of data to the platform, and construct a full-blown model of advertising or dynamic pricing to derive the revenues generated by the data. Third, taxes may have other welfare effects which are not captured in our model. Taxes may affect the behavior of the platform, possibly leading the platform to exclude some users and advertisers or to start charging a subscription fee. This calls for extending the analysis. We plan to expand the analysis by allowing the platform to choose a larger menu of policies with different degrees of data exploitation in order to discriminate more finely among users.

7 Proofs

Proof of Lemma 1: By implicit differentiation of equation 4

$$
\begin{aligned}
\frac{\partial \xi}{\partial u_0} &= \frac{1}{1 - \alpha a \xi^{\alpha-1}} > 0, \\
\frac{\partial \xi}{\partial a} &= \frac{\xi^{\alpha}}{1 - \alpha a \xi^{\alpha-1}} > 0, \\
\frac{\partial \xi}{\partial \alpha} &= \frac{a \xi^{\alpha} \log \xi}{1 - \alpha a \xi^{\alpha-1}} < 0.
\end{aligned}
$$

∎

Proof of Proposition 1: The function $\Pi(x)$ is strictly increasing in x for $x \leq \xi(u_0)$ and for $x > \xi(u_0)$, we show that

$$\frac{x\Pi'(x)}{\Pi(x)} = \frac{\beta v(x)}{v_0 + v(x)} + \frac{\alpha u(x)}{u_0 + u(x)} - 1,$$

is non-decreasing in x and strictly increasing if $u_0 > 0$ or $v_0 > 0$. But this implies that there cannot be an interior maximum of $\Pi(x)$ in $(\xi(u_0), 1)$ as this would imply that Π' switches from a positive to a negative value, contradicting the fact that $\frac{x\Pi'(x)}{\Pi(x)}$ is non-decreasing in x. We now compare $\Pi(\xi(u_0))$ and $\Pi(1)$. We have

$$\Pi(\xi(u_0)) = v_0 + b\xi(u_0)^\beta \text{ and } \Pi(1) = (v_0 + b)(u_0 + a) \tag{5}$$

showing that $\xi(u_0)$ is chosen when $\frac{v_0}{b} \leq \nu$.

We compute the comparative statics effects of changes in the parameters on ν.

$$\frac{\partial \nu}{\partial \alpha} = -\frac{\partial \xi(u_0)}{\partial \alpha} > 0,$$
$$\frac{\partial \nu}{\partial \beta} = -\log x x^\beta > 0.$$

To compute the effect of changes in u_0, we compute the sign of

$$\frac{\partial \nu}{\partial u_0} = 1 - \beta\xi(u_0)^{\beta-1}\frac{\partial \xi}{\partial u_0}(1 - u_0 - a) - \xi(u_0)^\beta.$$

When u_0 converges to $1 - a$, $\frac{\partial \nu}{\partial u_0}$ converges to zero. We also compute the second derivative of ν with respect to u_0:

$$
\begin{aligned}
\frac{\partial^2 \nu}{\partial u_0^2} &= \beta\xi(u_0)^{\beta-1}\frac{\partial \xi}{\partial u_0} - \beta(\beta-1)\xi(u_0)^{\beta-2}[\frac{\partial \xi}{\partial u_0}]^2(1 - u_0 - a) \\
&- \beta\xi(u_0)^{\beta-1}\frac{\partial^2 \xi}{\partial u_0^2}(1 - u(0) - a) - \beta\xi(u_0)^{\beta-1}\frac{\partial \xi}{\partial u_0} \\
&= -\beta(\beta-1)\xi(u_0)^{\beta-2}[\frac{\partial \xi}{\partial u_0}]^2(1 - u_0 - a) - \beta\xi(u_0)^{\beta-1}\frac{\partial^2 \xi}{\partial u_0^2}(1 - u(0) - a)
\end{aligned}
$$

Next note that

$$\frac{\partial^2 \xi}{\partial u_0^2} = \frac{a\alpha(\alpha-1)\xi(u_0)^{\alpha-2}}{(1-a\alpha x^{\alpha-1})^2} \frac{\partial \xi}{\partial u_0} < 0,$$

so that we finally obtain $\frac{\partial^2 \nu}{\partial u_0^2} > 0$ which together with the fact that $\frac{\partial \nu}{\partial u_0} = 0$ at $u_0 = 1 - a$ guarantees that $\frac{\partial \nu}{\partial u_0} < 0$. ∎

Proof of Proposition 2: The welfare admits two different expressions for $x \leq \xi(u_0)$ and $x \geq \xi(u_0)$. We observe that the derivative $W'(x)$ is continuous at $\xi(u_0)$, is continuous, by comouting the left-hand and right-hand derivatives, $W'^{-}(\xi(u_0))$ and $W'^{+}(\xi(u_0))$:

$$
\begin{aligned}
W'^{-}(\xi(u_0)) &= u'(\xi(u_0)) - \frac{1}{2}, \\
W'^{+}(\xi(u_0)) &= \frac{[u_0 + u(\xi(u_0))][2\xi(u_0)u'(\xi(u_0)) - (u_0 + u(\xi(u_0)))]}{2\xi(u_0)^2} \\
&= u'(\xi(u_0)) - \frac{1}{2},
\end{aligned}
$$

where the last equality stems from the fact that $\xi(u_0) = u_0 + u(\xi(u_0))$. This implies that generically, $\xi(u_0)$ will never be the optimal level of data exploitation. When $x \leq \xi(u_0)$, all users access the platform, and welfare is strictly concave in x. The maximum in this region is attained at $\hat{x} = (2a\alpha)^{\frac{1}{1-\alpha}}$ if $\hat{x} \leq \xi(u_0)$ and at $\xi(u_0)$ otherwise. In the latter case, the derivative is positive at $\xi(u_0)$ and $\xi(u_0)$ cannot be the global optimum.

When $x | geq \xi(u_0)$, we compute

$$\frac{xW'(x)}{W(x)} = \frac{\alpha u(x)}{u_0 + u(x)} - 1,$$

The elasticity of welfare is non-decreasing in x, so that the maximum in the exclusion region is either attained at $\xi(u_0)$ (if the derivative is negative at that point) or at 1. We conclude that the global optimum is either achieved at $\hat{x}$ or at 1.

To determine more precisely the optimum, we consider the simple case where the access benefit is zero, $u_0 = 0$. Since $\hat{x} = (2a\alpha)^{\frac{1}{1-\alpha}}$ and $\xi(0) = (a)^{\frac{1}{1-\alpha}}$, $\hat{x} < \xi(0)$ if and only if $\alpha < 1/2$. Furthermore, the elasticity of welfare is constant equal to $2\alpha - 1$ in the exclusion region, so that welfare decreases if $\alpha < \frac{1}{2}$ and increases if $\alpha > \frac{1}{2}$. We conclude that, when $u_0 = 0$, welfare is

maximized at $\widehat{x}$ if $\alpha \leq \frac{1}{2}$ and at 1 if $\alpha \geq \frac{1}{2}$.

Now, $W(\widehat{x}) - W(1) = u_0 + u(x) - \frac{(a+u_0)^2}{2}$ so that $\frac{\partial W(\widehat{x}) - W(1)}{\partial u_0} = 1 - (a + u_0) > 0$. As u_0 increases, the welfare under market coverage increases faster than under exclusion, as every marginal increase in u_0 raises welfare by 1 under market coverage but by $a + u_0 < 1$ under exclusion. Hence, when u_0 increases, users are more likely to prefer market coverage at $\widehat{x}$. When u_0 is equal to $1 - a$, one checks that $W(\widehat{x}) > W(1)$ when $\alpha > \frac{1}{2}$, completing the proof of the Proposition. ∎

Proof of Proposition 3: The Proposition is a simple consequence of the comparative statics results of Lemma 1 and Proposition 1, we observe that a reduction in v_0 increases the region of parameters for which exclusion is chosen while not affecting the coverage level $\xi(u_0)$. A reduction of $v(x)$, which can be interpreted as a reduction in b, leaves $\xi(u_0)$ unaffected and reduces the region of parameters for which exclusion is chosen. A reduction in u_0 increases the bound ν, thereby increasing the region of parameters for which exclusion occurs. However, a reduction in u_0 also reduces the coverage level, thereby reducing the degree of data exploitation when the platform chooses to cover the market. ∎

Proof of Proposition 7: If $\alpha + \beta \leq 1$, the platform chooses coverage at $\xi(0) < \xi(u_0)$ under the binary policy. As $\alpha < \frac{1}{2}$, $\widehat{x} \leq \xi(0) < \xi(u_0)$, and hence users favor a lower value of data collection and prefer the régime with the opt-out option. If now $\alpha + \beta > 1$ and $u_0 < \overline{u_0}$, the platform selects the maximal level of data collection under both régimes and users prefer the opt-out option as it allows excluded users to collect the access benefit u_0. Finally, consider the case where $\alpha + \beta > 1$ and $u_0 \geq \overline{u_0}$. In the baseline scenario, the platform covers the market and welfare is

$$W = u_0 + a\xi(u_0)^{\alpha} - \frac{\xi(u_0)}{2} = \frac{xi(u_0)}{2},$$

With the outside option, the platform chooses the maximal degree of data exploitation and welfare becomes

$$W' = u_0 + \frac{a^2}{2},$$

. The conclusion follows. ∎

8 Bibliography

ANDERSON,S. AND S. COATE (2005) "Market provision of broadcasting: A welfare analysis ," *Review of Economic Studies*, **72**, 947-972.

ANDERSON,S. AND J. GABSZEWICZ (2006) "The media and advertising: A tale of two-sided markets," in *Handbook of Cultural Economics*, (V. Ginsburgh and D. Throsby (eds.)), North Holland.

ANDERSON,S. AND J. GANS (2011) "Platform siphoning: Ad-avoidance and media content ," *American Economic Journal: Microeconomics* , **3**, 1-34.

BOURREAU, M., B. CAILLAUD AND R. DE NIJS (2015) "Digital platforms, advertising and taxation ," mimeo., PSE Mines Telecom.

CASADESUS-MASANELL, R. AND A. HERVAS-DRANE (2015) "Competing with privacy" *Management Science*, **61**, 229-246.

CHOI, J.P. (2006) "Broadcast competition and advertising with free entry: Subscription vs. free-to-air," *Information economics and Policy*, **18**, 181-196.

CRAMPES, C., C. HARITCHABALET AND B. JULLIEN (2009) "Advertising, competition and entry in media industries ," *Journal of Industrial Economics* , **57**, 7-31.

CRÉMER, J. (2015) "Taxing network externalities ," mimeo., PSE Toulouse.

DE CORNIÈRE, A. AND DE NIJS R. (2016) "Online advertising and privacy ," *RAND Journal of Economics*, **47**, 48-72.

GABSZEWICZ, J.J., D. LAUSSEL AND N. SONNAC (2001) "Press advertising and the ascent of the pensee unique ," *European Economic Review* , **45**, 645-651.

GABSZEWICZ, J.J., D. LAUSSEL AND N. SONNAC (2004) "Programming and advertising competition in the broadcasting industry ," *Journal of Economics and Management Strategy*, **13**, 657-669.

JOHNSON, J. (2013) "Targeted advertising and advertising avoidance ," *RAND Journal of Economics*, **44**, 128-144.

KIND, H.J. AND SCHJELDERUP, G. AND STÄHLER, F. (2013) "Newspaper Differentiation and Investments in Journalism: The Role of Tax Policy," *Economica*, **80**, 131-148.

KIND, H.J. AND KOETHENBUERGER, M. AND SCHJELDERUP, G. (2010A) "Tax responses in platform industries," *Oxford Economic Papers*, **62**, 764-783.

KIND, H.J. AND KOETHENBUERGER, M. AND SCHJELDERUP, G. (2010B) "On revenue and welfare dominance of ad valorem taxes in two-sided markets," *Economics Letters*, **104**, 86-88.

KIND, H.J. AND KOETHENBUERGER, M. AND SCHJELDERUP, G. (2008) "Efficiency enhancing taxation in two-sided markets," *Journal of Public Economics*, **92**, 1531-1539.

KOTSOGIANNIS, C. AND SERFES, K. (2010) "Public goods and tax competition in a two-sided market," *Journal of Public Economic Theory*, **12**, 281-321.

PEITZ, M. AND T. VALLETTI (2008) "Content and advertising in the media: Pay-tv vs free-to-air" *International Journal of Industrial Organization*, **26**, 949-965.

REISINGER, M., L. RESSNER AND R. SCHMIDTKE (2009) "Two-sided markets with pecuniary and participation externatlities" *Journal of Industrial Economics*, **57**, 32-57.

REISINGER, M. (2012)) "Platform competition for advertisers and users in media markets" *International Journal of Industrial Organization*, **30**, 243-252.

TAG, J. (2009) "Paying to remove advertisements," *Information Economics and Policy*, **21**, 245-252.