Privacy and Platform Competition

Philipp Dimakopoulos (Humboldt University Berlin)
Slobodan Sudaric (Humboldt University Berlin)

Discussion Paper No. 67

January 22, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Privacy and Platform Competition

Philipp Dimakopoulos
Humboldt-Universität zu Berlin[*]

Slobodan Sudaric
Humboldt-Universität zu Berlin[†]

January 8, 2018

Abstract

We analyze platform competition where user data is collected to improve ad-targeting. Considering that users incur privacy costs, we show that the equilibrium level of data provision is distorted and can be inefficiently high or low: if overall competition is weak or if targeting benefits are low, too much private data is collected, and vice-versa. Further, we find that softer competition on either market side leads to more data collection, which implies substitutability between competition policy measures on both market sides. Moreover, if platforms engage in two-sided pricing, data provision is efficient.

JEL–classification numbers: D43, L13, L40, L86

Keywords: ad targeting, platform competition, privacy, user data

[*]HU Berlin, Spandauer Str. 1, 10178 Berlin, Germany, Email: philipp.dimakopoulos@hu-berlin.de

[†]HU Berlin, Spandauer Str. 1, 10178 Berlin, Germany, Email: slobodan.sudaric@hu-berlin.de

The authors would like to thank Pio Baake, Helmut Bester, Colin von Negenborn, Markus Reisinger, Roland Strausz, Jonas von Wangenheim as well as the guest editor, Thomas Ross, and two anonymous refeeres for helpful comments and suggestions. Further we thank participants of CRESSE on Crete, IODE in Liège, MaCCI in Mannheim, BCCP in Berlin, EARIE in Maastricht, and seminar participants in Berlin. All remaining errors are of course our own. Financial support from the Deutsche Forschungsgemeinschaft (RTG 1659 and CRC 190) is gratefully acknowledged. Note that a previous version of this paper circulated under the title "User Data and Platform Competition".

1 Introduction

Online platforms often do not charge monetary prices from users but monetize through an advertisement-based business model building on the collection and processing of user data. Typical examples include social networks (e.g. Facebook, LinkedIn), search engines (e.g. Bing, Google) or video platforms (e.g. Youtube, Vimeo). The role of user data in this context is ambiguous. From the platform perspective user data is an input factor which can be used to gain insights about users and improve the targeting of advertisement, resulting in a superior product for potential advertisers. This commodity attribute of data is mirrored to a lesser extent on the user side. Users typically accept some conditions to what extent personal data is collected and processed when using a platform service. In some cases the provision of personal data is necessary to make meaningful use of a platform service (e.g. social networks) while in other cases services do not require the collection of user data per se (e.g. search engines, mail providers, video platforms). In both cases the provision of data from a user perspective can be interpreted as a price the user is willing to accept in exchange for the use of the platform including the display of ads.[1] To put it in terms of platform economics, user data requirements exhibit price characteristics on the one hand, and affect indirect network effects (e.g. targeting) at the same time.

This ambiguity makes it especially hard for policy makers as standard economic arguments might not be applicable. Indeed, the *European Data Protection Supervisor (EDPS)* argues that competition authorities should take privacy and data related aspects more into account (EDPS, 2014).[2] And indeed, recent cases demonstrate that competition authorities acknowledge the peculiarities of data-driven industries. Germany's *Federal Cartel Office (Bundeskartellamt, BKartA)* initiated investigations against Facebook in 2016 based on alleged abuse of market power. In particular, the BKartA investigates whether Facebook uses its dominant position in the market for social networks to expand the terms of service

[1] A study by the Pew Research Center (2014) shows that 91 percent of respondents agree that they lost control over how companies collect personal data while 55 percent state that they are willing to share some information in exchange for using a free service. The European Commission (2015), however, reports that 72 percent of internet users worry that they provide too much data online. This indicates that users are aware and willing to exchange personal data for services, however, the actual extent worries them.

[2] Whether competition authorities should incorporate aspects of privacy and data protection is, however, controversial. For arguments in favor we refer to Stucke and Grunes (2016), arguments against can be found e.g. in Cooper (2013).

outlining how much data is collected and processed by the platform.[3] Therefore, we want to shed some light on the role of competition intensity in a two-sided market framework when users provide data and this data is monetized on the opposing market side.[4]

We analyze a setting of two competing ad-financed platforms in a two-sided market framework. On the user market side, platforms strategically set the required level of data provision, to which users have to agree to obtain access to the platform service. Platforms process this user data to sell improved ad targeting on the advertiser market side. While users incur disutility from providing data (privacy concerns, opportunity costs), they benefit from seeing more relevant ads. Users and advertisers are assumed to single-home.

Our model predicts that platforms will extract a distorted amount of data compared to the efficient benchmark. The distortion is induced through the one-sided monetization in a way that platforms do not perfectly balance the costs of data provision, i.e. privacy costs incurred by users, against the targeting benefits on both market sides, but put too much or too little weight on the benefit captured by the monetized market side. This distortion depends on the net effect of cross-group externalities as well as the degree of competition intensity on both market sides. If targeting benefits are small or competition is weak, an inefficiently high level of data is collected. On the other hand, if competition is strong or targeting benefits sufficiently outweigh nuisance costs, too little data is collected. From the point of view of consumers the competitive level of data provision is always too high, suggesting that applying a consumer standard to online platforms leads to underprovision of personal data. The competitive equilibrium level of data provision, however, is monotone in the degree of competition intensity: the weaker the competition on either side of the market, the higher the equilibrium amount of data provision. This result is interesting because it does not follow the common two-sided platform logic that less elasticity on one side typically decreases the other side's price.

[3] Bundeskartellamt, 'Bundeskartellamt initiates proceeding against Facebook on suspicion of having abused its market power by infringing data protection rules', Press Release, 2 March 2016, http://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2016/02_03_2016_Facebook.html.

[4] Classical examples include ad-based business models where data is used to improve ad targeting or matching / recommendation platforms, where users are presented offers which become more relevant the more the platform knows about its users. For illustration purposes we stick to the example of targeted advertising and refer to the extension part of this paper for a more general consideration of cross-group externalities, i.e. also the possibility of users enjoying the presence of firm's offers.

Our findings indicate that the inefficiency of data provision can be reduced by careful privacy regulation or competition policies on either market side. One interpretation of this result is that (competition) policy measures in these data-driven industries should take into account the effects they have on the extent of private data collection.

We also consider a variety of extensions to this setup. In the first one we depart from the assumption that platforms are restricted in their price setting on the user side, and allow for non-zero user prices. In fact, lifting the restriction leads to an efficient level of collected data, while user prices can be positive, negative (or zero). This gives rise to two interpretations. The first is a Coasian one, where establishing the missing market on the user side leads to an efficient outcome. This reflects the idea of Laudon (1996) that users should be adequately compensated for the provision of their data, while the problem of the 'data economy' lies precisely in the absence of such a market. The second interpretation is of counterfactual nature. In particular we argue that whenever the unrestricted model would yield positive (negative) user prices, the restricted model exhibits overprovision (underprovision) of user data as platforms can no longer adequately charge or compensate users for collecting data. The second extension considers different degrees of platform collusion and we conclude that the amount of collected data is excessively high under full collusion, while this is not necessarily the case under partial collusion. In the third extension we discuss the robustness of our results with respect to multi-homing and elastic total demand. Lastly, we demonstrate that our results naturally extend to settings with positive cross-group externalities (matching platforms).

The remaining paper is structured as follows. Section 2 relates our analysis to the existing literature. Section 3 introduces the model. Section 4 characterizes the efficient benchmark and competitive equilibrium outcomes, for which we present comparative statics in Section 5. Section 6 compares these outcomes and outlines policy implications. In Section 7 we extend and discuss the baseline model. Section 8 concludes.

2 Related literature

Methodologically, our research is related to the literature on platform competition in general and on applications in media markets in particular. We consider a competitive setting with two-sided single-homing which has been analyzed by Armstrong (2006) in a more general framework and later extended in Armstrong and Wright (2007). However, both papers consider the case where platforms engage in two-sided pricing while non-monetary aspects (as e.g. user data) are not modelled. We also share a common component with the literature on media platforms in the sense that we, at least in our baseline model, consider the case of opposing indirect network effects, where advertisers like to reach many users but users dislike the presence of advertisers. This reflects the idea of 'peace and quiet' privacy in Posner (1981) and is a common assumption in the media literature (see Anderson and Gabszewicz (2006) for a review). This setup is used e.g. to study competition in TV markets (see e.g. Anderson and Coate (2005) or Peitz and Valletti (2008)) where platforms do not engage in targeted advertising and therefore the expected revenue per user as well as perceived nuisance are constant. Our research differs in the sense that we endogenize those indirect network effects as we let them to be affected by the level of data collected. The concept of endogenous network effects is captured in Reisinger (2012) where users spend time using platform services and platforms translate this activity into better targeting and reduced nuisance. A similar setup is presented in Bourreau et al. (2017), however the research question differs substantially. The key difference is that in our model the level of data provision is a strategic decision of the competing platforms, while in the two previously mentioned papers consumers voluntarily spend time/provide data on the platforms which changes the competitive dynamics significantly.

We also contribute to the broader literature on efficient provision of personal data and the role of privacy as a competition instrument. The aspect of data provision being a strategic choice made by platforms is captured to some extent by Spiegel (2013) who compares commercial software (full privacy) to adware (positive privacy costs) and shows that adware is welfare superior. De Corniere and De Nijs (2016) consider a setting where a monopolistic platform auctions off advertising slots and decides whether to disclose consumer information (no privacy) or not (privacy). They show that platforms might prefer information disclosure, which comes at the cost of some consumers leaving the market such that from a welfare point of view it is not clear which regime is preferable. Bloch and Demange (2017) present a setting where consumers are heterogeneous with respect to their privacy cost and a monopolistic platform decides how much data to extract. They show that depending on parameter values the amount of data collection can be excessively high. A similar setting is presented in Lefouili and Toh (2017) where a monopolistic platform monetizes on disclosing personal information to third parties. The authors conclude that one of the inefficiencies arising is excessive information disclosure. The mentioned papers consider the case of monopolistic platforms, while we consider the case of competing platforms, allowing for varying degrees of competition intensity on both market sides.

The role of privacy in a competitive environment is considered in Casadesus-Masanell and Hervas-Drane (2015) where firms not only compete in a price dimension but also in a quality dimension which the authors motivate as privacy. They show that compared to a monopolistic firm, competition leads to a higher degree of privacy while increasing competition intensity does not necessarily imply that privacy improves even further. A key assumption in their model is that prices for disclosing consumer information are exogenous, while in our model platforms have market power vis-à-vis advertisers and hence face a tradeoff. They also show that low privacy firms tend to subsidize consumers, while high privacy firms charge positive consumer prices. Similarly, Kummer and Schulte (2016) show empirically that there is a trade-off between money and privacy for users. They analyze mobile application data and find that apps are cheaper when more personal data can be collected. These results reoccur in our two-sided pricing extension as we show that user prices can be positive or negative as well, while the degree of privacy provision is excessively high or low once firms can no longer compensate users for their data provision. To our knowledge there are very few empirical studies examining the interaction between market power and privacy. In fact, the only study we are aware of is Bonneau and Preibusch (2010) who relate the extent of data collection policies of various online services to the competitiveness of the market they are operating in. They show that the more market power a firm has, the more personal information is asked to be provided which is in line with our model.

3 Model

We analyze a setting where two symmetric platforms, $i, j \in \{1, 2\}$ with $j \neq i$, compete for advertisers and users. Advertisers and users are distributed uniformly on different Hotelling lines of unit length and are assumed to both single-home. This assumption allows us to focus on the role of competition intensity more clearly.[5] Platforms are located at the ends of the respective Hotelling lines such that platform i is located at location $l_i = 0$ and platform j at $l_j = 1$. Note that on the advertiser and the user side we have distinct Hotelling lines and therefore distinct parameters of transportation costs, which we will later interpret as different degrees of competition intensity. The idea is that the degree of competition faced by platforms does not have to be same for all market sides. For example, online platforms from different segments, such as search engines, social networks, video streaming platforms or mail providers, may all compete for the same advertisers, however competition for users may occur separately and independently of the other segments.

3.1 Users

A user located at x on the Hotelling line obtains utility $u_i(x)$ from joining platform i,

$$u_i(x) = \underline{u} - \kappa(d_i) - \nu(d_i)A_i - t_u|l_i - x|. \tag{1}$$

The first term of the utility function is a fixed utility component $\underline{u}$ from using platform services, which is the same for both platforms. Second, $\kappa(d_i) \geq 0$ denotes the privacy (opportunity) costs of providing user data d_i to the platform, whereby we assume that costs are strictly convex and twice differentiable, and specifically that $\kappa'(0) = 0$, while $\kappa'(d) > 0$ for all $d > 0$ and $\kappa''(d) > 0$ for all d. Third, users incur nuisance cost $\nu(d) \geq 0$ per advertisements A_i on the platform. We assume that users (weakly) prefer personalized to non-personalized ads, i.e. $\nu(d)$ is a convex and twice differentiable function s.t. $\nu'(d) \leq 0$ and $\nu''(d) \geq 0$. This setup reflects the idea that the more relevant an ad, the higher the chance of value creation through a possible follow-up purchase.[6] Finally, users face transportation costs due to horizontal platform differentiation, whereby we assume uniform user distribution on the Hotelling line, i.e. $x \overset{u}{\sim} [0, 1]$, while $t_u > 0$ is the associated transportation cost parameter.

[5] In Section 7 we discuss multi-homing.

[6] Note that our set-up allows for positive utility of seeing advertisement as well, as long as this positive utility is again concave in the amount of provided data. However, for sake of clarity we stay with the notion of negative utility of nuisance in the subsequent text and consider the case of positive cross-group externalities as an extension in Section 7.

Consumers in our baseline model are not charged a monetary price explicitly, which makes our model comparable to e.g. Reisinger (2012). We follow the same line of reasoning as e.g. in Peitz and Reisinger (2016) and Waehrer (2015) that there are some exogenous constraints preventing platforms from charging non-zero consumer prices. This restriction is, however, relaxed in Section 7.1. In order to join a platform users have to provide some personal data d_i in our model. This is different to the setup in Reisinger (2012) or Bourreau et al. (2017) as in our model platforms can set the level of data which has to be provided by the users, whereas in their models consumers voluntarily provide a certain amount of time. The idea behind our setup is that consumers accept terms and conditions when using a platform which requires them to accept a certain level of data provision or alternatively cases where users have to register for an account by providing personal information before they can use the platform service. This specification on the consumer side allows us to focus on user data d_i as primary strategic aspect for competition.

3.2 Advertisers

An advertiser located at a on the Hotelling line obtains an expected profit of $\pi_i(a)$ from posting a single ad on platform i,

$$\pi_i(a) = \tau(d_i)(1 - p_i)X_i - t_a|l_i - a|. \tag{2}$$

The interaction with X_i users on platform i generates a normalized expected revenue of 1, if users decide to 'click on the ad', which happens with probability $\tau(d_i)$. The strictly concave and twice differentiable function $\tau(d) \geq 0$ can be interpreted as the targeting ability of platforms: the more data d can be collected from users, the more effective the targeting and hence the higher the probability that a user clicks on this ad, i.e. we have that $\tau'(d) > 0$ and $\tau''(d) < 0$. At the same time we assume that advertisers only pay the platform a price p_i if the ad has been clicked (cost-per-click) such that the expected revenue per user is given by $\tau(d_i)(1 - p_i)$, which is consistent with real-world pricing practices. The second term reflects advertisers transportation costs when joining platform i. Again we assume uniform advertiser distribution on the Hotelling line, i.e. $a \overset{u}{\sim} [0, 1]$, and $t_a > 0$ as the transportation cost parameter on the advertiser side.

3.3 Platforms

The business model of platforms in our model is purely ad-based. While they offer (exogenous) platform services ($\underline{u}$) to users, revenue is only generated through presenting ads to users.[7] Platform profits are then given by

$$\Pi_i \left(d_i, p_i\right) = A_i X_i \tau(d_i) p_i \tag{3}$$

i.e. A_i advertisers at platform i pay p_i whenever the platform's users X_i click on an ad with probability $\tau(d_i)$.[8] The crucial novelty in our model is that we assume that besides charging prices to advertisers, platforms extract data d_i from their users. While d_i shares some price characteristics from the point of view of users, data is an essential input factor for the click-probability the advertisers are facing. At the same time we assume that not only the click probability increases through better targeting possibilities but also the nuisance decreases.

3.4 Assumptions

We make the following assumptions to ensure full advertiser and user market coverage, allowing us to study environments of full platform competition.[9]

Assumption 1 *Competition for advertisers is sufficiently strong, i.e.* $t_a \leq \bar{t}_a$.
This implies that competition for users is sufficiently weak and that there are gains of trade for all advertisers, even without data collection, i.e.

$$(a) \quad t_u > \nu(0),$$
$$(b) \quad t_a < \tau(0).$$

The upper bound on t_a is given by $\bar{t}_a := \frac{t_u \tau(0) - \nu(0)\tau(0)}{3t_u + \nu(0)}$. This assumption on the upper bound of t_a allows us to isolate effects in a competitive environment. Intuitively, this

[7] In Section 7 we discuss two-sided pricing.

[8] Note that platforms and advertisers share the profit created by each targeted user on the platform. However, this does not mean that their incentives are perfectly aligned, since platforms additionally care about the number of advertisers joining.

[9] In Section 7 we discuss relaxing the full-market coverage assumptions.

constitutes a sufficient condition, such that for any level of (symmetric) data provision $d \geq 0$, it is assured that all advertisers obtain non-negative profits. Consequently, competition for advertisers is sufficiently strong.

The condition on the consumer nuisance function, i.e. the necessary condition (a) of Assumption 1, can be motivated as follows: no platform will obtain the entire user market, even if all ads were placed on the rival platform. Technically, this is established by $t_u > \nu(0)$.[10] The condition on the targeting technology, i.e. the necessary condition (b) of Assumption 1, states that even without collecting any data advertisers can still profitably join a platform. In particular we assume that there are gains of trade for all advertisers. Intuitively, this assumption states that there is a positive probability for users to click an ad even if the ad is not targeted at all. And this probability, $\tau(0)$, exceeds the transportation cost incurred by any advertiser t_a, so that we need not exclude any advertisers, even if too little data is collected.

Assumption 2 *The fixed utility component $\underline{u}$ is large enough to ensure full participation on the user side.*

Intuitively, the platform service provides sufficient utility such that users are not deterred through the provision of personal data and seeing ads.

The timing of the game is as follows. In the first stage platforms simultaneously set prices p_i and the required level of data d_i to join their platform. In the second stage advertisers and users observe the platforms' choices and simultaneously decide which platform to join, hence determining A_i and X_i.[11] The equilibrium concept is subgame perfection and we solve the game by backward induction.

4 Equilibrium analysis

In this section we will first present the results for the second-stage subgame of user and advertiser allocation. Then we will show the efficient and the user-optimal outcome as well

[10]Note that $t_u > \nu(0) \Rightarrow t_u > \nu(d) \, \forall d$ because $\nu'(d) \leq 0$. Given any (symmetric) amount of data $d \geq 0$ collected by both platforms, even if all advertisers used platform j such that $A_i = 0$ and $A_i = 1$, at least the user most loyal to platform j, i.e. located directly at l_j, would rather stay at this platform j, even though it is full of ads. In other words, competition for users is sufficiently weak.

[11]We could also consider an alternative timing where advertisers choose first and users last. The outcome is equivalent in our model.

as the market outcome in the Subgame Perfect Nash Equilibrium.

4.1 Second stage market shares

In the second stage the market shares on the consumer and advertiser side are given by the standard Hotelling procedure. Utilizing the unit length of the Hotelling line, and given full user market coverage due to Assumption 2, the number of users joining a platform is then determined by the indifferent consumer $\hat{x} : u_i(\hat{x}) = u_j(\hat{x})$ such that

$$X_i = \hat{x} = \frac{1}{2} + \frac{1}{2t_u}\left[\kappa(d_j) - \kappa(d_i) + \nu(d_j)A_j - \nu(d_i)A_i\right], \qquad X_j = 1 - \hat{x}. \tag{4}$$

Similarly, market shares on the advertiser side are given by the indifferent advertiser $\hat{a} : \pi_i(\hat{a}) = \pi_j(\hat{a})$. Note that Assumption 1 assures market coverage gross of advertising prices. For now we therefore assume that prices permit full market coverage and check later that in equilibrium this is indeed the case. Market shares are then given by

$$A_i = \hat{a} = \frac{1}{2} + \frac{1}{2t_a}\left[\tau(d_i)(1-p_i)X_i - \tau(d_j)(1-p_j)X_j\right], \qquad A_j = 1 - \hat{a}. \tag{5}$$

Solving the system of equations given in (4) - (5) yields unique market shares X_i, X_j, A_i and A_j as functions of data requirements d_i, d_j and prices p_i, p_j. Explicit solutions are provided in the Appendix.

4.2 Efficiency benchmark

For the derivation of the welfare-efficient benchmark, we define welfare as the sum of all indirect utilities and profits, anticipating second stage market shares as in 4.1, i.e.

$$W(d_i, d_j, p_i, p_j) = \int_0^{X_i} u_i(x)\mathrm{d}x + \int_{X_i}^1 u_j(x)\mathrm{d}x + \int_0^{A_i} \pi_i(a)\mathrm{d}a + \int_{A_i}^1 \pi_j(a)\mathrm{d}a + \Pi_i + \Pi_j. \tag{6}$$

Proposition 1 *Welfare is maximized by the unique symmetric solution* $(d^o, p^o) = (d_i^o, p_i^o)$

for $i \in \{1, 2\}$, where d^o is characterized by

$$\kappa'(d^o) = \frac{\tau'(d^o)}{2} - \frac{\nu'(d^o)}{2} \tag{7}$$

resulting in equal advertiser and user market shares, i.e. $A_i^o = 1/2$ and $X_i^o = 1/2$. The price p^o can be freely chosen to split the rent between advertisers and platforms.[12]

The welfare-optimal level of data d^o is chosen in a way such that users' marginal cost of data provision $\kappa'(d^o)$ equals the sum of marginal benefits across both market sides, i.e. the marginal benefit of enhanced targeting $\tau'(d^o)/2$ and the marginal benefit of reduced nuisance $-\nu'(d^o)/2$, while the factor 1/2 is due to the symmetric market shares.[13] Furthermore, the optimal level of data provision is independent of transportation cost parameters t_a and t_u. Since prices are just transfers from advertisers to platforms they do not affect welfare.[14]

4.3 User-optimal outcome

Let us now turn to the user-optimal level of data provision. If users are free to decide on the amount of data provided, the user-optimal level d^u is derived from consumer surplus, which is identical to the first two terms in equation (6), anticipating second stage market shares as in 4.1.[15]

Proposition 2 *User utility is maximized by the unique symmetric solution $(d^u, p^u) =$*

[12]Note that p^o has to be sufficiently small such that the advertiser market remains fully covered. The upper bound on p^o is then obtained from the participation constraint of the indifferent advertiser at $a = 1/2$ such that $\pi_i(1/2) \geq 0 \iff p^o \leq 1 - t_a/\tau(d^o) < 1$. The last inequality follows from Assumption 1.

[13]For very low transportation cost parameters and sufficiently high net benefits $\tau(\cdot) - \nu(\cdot)$ on the platform it might be efficient from a welfare perspective to shut one platform down and let the entire market be served by the other platform due to high network effects. In this case the very fact of having a competing platform is an inefficiency. While this corner solution exhibits an interesting property of platform markets, it is not the focus of this paper and we therefore stick to the case where we have an interior, i.e. duopoly solution as the efficient benchmark.

[14]The same data level d^o would result if we only choose d_i to maximize welfare, while anticipating firms setting ad prices p_i subsequently. These prices would be identical to the prices in the market outcome, given by equation (13). The same argument applies for the user optimal level d^u.

[15]See footnote 14.

(d_i^u, p_i^u) *for* $i \in \{1,2\}$*, where* d^u *is characterized by*

$$\kappa'(d^u) = -\frac{1}{2}\,\nu'(d^u), \tag{8}$$

while the price p^u *can be freely chosen to split the rent between advertisers and platforms, resulting in equal advertiser and user market shares, i.e.* $A_i^u = 1/2$ *and* $X_i^u = 1/2$*.* [16]

Intuitively, the user-optimal data level balances privacy costs and reduced nuisance benefits for users, at the margin. Note that for constant nuisance costs we get the corner-solution where users would not provide any private data, i.e. $d^u = 0$. For general decreasing nuisance costs, users would be willing to provide a positive level of data $d^u > 0$.

4.4 Market outcome

For the market outcome, in the first stage platforms maximize their profits, anticipating second stage market shares as in Section 4.1.

$$\max_{p_i, d_i} \Pi_i\,(d_i, p_i) = A_i\,\tau(d_i)\,p_i\,X_i \;\forall i \in \{1,2\} \tag{9}$$

We obtain solutions for prices and data levels from the first-order conditions, i.e.

$$\frac{\tau'(d_i)}{\tau(d_i)} = -\frac{\frac{\partial A_i}{\partial d_i} X_i + \frac{\partial X_i}{\partial d_i} A_i}{A_i\,X_i}, \tag{10}$$

$$p_i = \frac{A_i\,X_i}{\frac{\partial A_i}{\partial p_i} X_i + \frac{\partial X_i}{\partial p_i} A_i}. \tag{11}$$

Intuitively, targeting benefits of data collection must equal the effects on user and advertiser shares, at the margin. Similarly, also prices must reflect their impact on user and advertiser shares. Regarding the curvature of the maximization problem we note that the solution to the first-order conditions represents a maximum as long as the targeting technology $\tau(\cdot)$ is sufficiently concave, the nuisance cost $\nu(\cdot)$ is sufficiently convex, or both. The details of this condition are given in Appendix A.

---

[16] Note that p^u has to be sufficiently small such that the advertiser market remains fully covered. The upper bound on p^u can be obtained as outlined in footnote 12.

Proposition 3 *There exists a (symmetric) Subgame Perfect Nash Equilibrium with* $(d_i^*, p_i^*) = (d^*, p^*)$ *for* $i \in \{1, 2\}$*, such that the level of data collected from a users is implicitly given by*

$$\kappa'(d^*) = \left(\frac{\nu(d^*) + t_u}{\tau(d^*) - t_a}\right) \frac{\tau'(d^*)}{2} - \frac{\nu'(d^*)}{2} \tag{12}$$

and prices per advertisement are

$$p^* = 2\frac{t_a t_u + \nu(d^*)\tau(d^*)}{\tau(d^*)\left[t_u + \nu(d^*)\right]}, \tag{13}$$

resulting in equal advertiser and user market shares, i.e. $A_i^* = 1/2$ *and* $X_i^* = 1/2$*.*

Comparing the market level of data provision d^* in (12) to the efficient level d^o in (7) we see that the marginal targeting benefit $\frac{\tau'(d^*)}{2}$ is additionally weighted by $\frac{\nu(d^*)+t_u}{\tau(d^*)-t_a}$. This distortion is analyzed in detail in chapter 6. Note that the equilibrium price p^* does not exceed one and that profits are positive for all advertisers due to Assumption 1.[17]

Before we continue we state a corollary concerning the equilibrium effect of data provision on user utility.

Corollary 1 *In equilibrium,* $\kappa'(d^*) > -\nu'(d^*)/2$*.*

Proof. See Appendix A.4. ■

Intuitively, Corollary 1 implies that in equilibrium users' data provision is such that the (negative) privacy costs effect on user utility is larger than the (positive) effect of reduced nuisance. Consequently, in the market outcome too much personal data is provided compared to the user-optimal level.[18]

5 Comparative statics

In this section we want to provide economic intuition for the equilibrium results of our model. For this we will provide comparative statics, given changes in advertiser-side competition intensity t_a and user-side competition intensity t_u as well as nuisance $\nu(d)$ and

[17] In Appendix A we provide the details for this result.

[18] In Section 6 we provide a detailed comparison of the market outcome and the user-optimal outcome.

targeting $\tau(d)$ on equilibrium values of personal data provision d^*, ad-per-click price p^*, as well as platform profits Π_i^*, advertiser profits π_i^* and user utility u_i^*.

As most of the comparative statics effects are in line with standard intuition from two-sided platforms, we delegate these analyses to the Online Appendix B and refer to the table in Figure 1 for an overview of all derived comparative statics results. In this section we focus on the important and seemingly counter-intuitive effects of competition intensities of both market side.

Figure 1: Overview of comparative statics

z	$\mathrm{d}d^*/\mathrm{d}z$	$\mathrm{d}p^*/\mathrm{d}z$	$\mathrm{d}\Pi_i^*/\mathrm{d}z$	$\mathrm{d}\pi_i^*/\mathrm{d}z$	$\mathrm{d}u_i^*/\mathrm{d}z$
t_a	+	+	+	−	−
t_u	+	−	−	+	−
$\nu(d)$	+	+	+	−	−
$\tau(d)$	−	?	+	?	+

Note that we distinguish between the platform competition intensity on the user side and on the advertiser side. As platforms are horizontally differentiated vis-à-vis both market sides, competition intensity on each side can be measured through the corresponding transportation cost parameter: higher transportation costs mean higher platform differentiation and thus higher switching costs on this market side, which can be interpreted as more platform market power and hence lower competition intensity.

5.1 Advertiser-side competition

First, we consider the effects of advertiser-side competition on data collection. For this consider the platform's first-order condition in equation (10) and note that the data level choice depends on the effects of d_i on advertiser and user market shares A_i and X_i. Regarding market share reactions we obtain $\partial X_i/\partial d_i < 0$ and $\partial A_i/\partial d_i < 0$ at equilibrium values.[19] Intuitively, additional data provision d_i would shy away users X_i because marginal privacy costs are higher than marginal benefits of reduced nuisance (compare Corollary 1). Although more data provision increases targeting, overall, advertisers would still be repelled by additional data provision because of the detrimental effect on user market share at that platform.

[19] Note that derivations can be found in Appendix A.5.

In equilibrium, if competition for advertisers softens, i.e. transportation costs t_a increase, advertisers become 'more sticky', i.e. less sensitive to changes in data provision (and hence user demand) such that $\partial^2 A_i/(\partial d_i \partial t_a) > 0$. Contrary, users become more sensitive to data provision such that $\partial^2 X_i/(\partial d_i \partial t_a) < 0$. Overall, the former effect dominates the latter effect in magnitude. Consequently, and recalling $X_i^* = A_i^* = 1/2$, the right-hand-side of equation (10) decreases in t_a such that the equilibrium level of data provision must increase as the left-hand-side is falling in d_i, i.e.

$$\frac{\mathrm{d}d^*}{\mathrm{d}t_a} > 0. \tag{14}$$

This effect might seem counter-intuitive initially. However note that in equilibrium platforms balance the following trade-off for the data level. On the one hand, more data collection yields higher targeting rates, higher advertiser demand and in sum higher profits. On the other hand, collecting more data decreases user demand, which in turn repels advertisers and thus decreases platform profits. If competition for advertisers softens, the latter effect is dampened more than the former effect is strengthened. This yields a new balance of the trade-off, where more user data is collected.

While advertiser prices p^* rise in t_a (compare Online Appendix B), the effect on user data collection d^* does not follow 'standard' two-sided platform logic as here less competition for advertisers, i.e. less sensitive advertiser demand, *increases* users' data 'payment'. Therefore, users actually benefit from increased competition on the advertiser side, such that also $\mathrm{d}u_i^*/\mathrm{d}t_a < 0$, as discussed in the Online Appendix B. Also, since $\mathrm{d}d^*/\mathrm{d}t_a > 0$ and $\mathrm{d}p^*/\mathrm{d}t_a > 0$ we naturally have $\mathrm{d}\Pi_i^*/\mathrm{d}t_a > 0$.

5.2 User-side competition

Second, we evaluate the effects of user-side competition intensity on data collection. Similar to the analysis above, we know that $\partial X_i/\partial d_i < 0$ and $\partial A_i/\partial d_i < 0$ in equilibrium. If competition for users softens, i.e. transportation costs t_u increase, on the one side users become less sensitive to changes in data provision such that $\partial^2 X_i/(\partial d_i \partial t_u) > 0$. Therefore, advertisers also become less sensitive to data provision such that $\partial^2 A_i/(\partial d_i \partial t_u) > 0$ because they care about the share of users on that platforms. Therefore the right-hand-side of equation (10) decreases in t_u such that the equilibrium level of data provision must increase, i.e.

$$\frac{\mathrm{d}d^*}{\mathrm{d}t_u} > 0. \tag{15}$$

Two effects are intuitively relevant here. On the one hand, platforms care about the share of users on their platform because it increases their profits directly, but also indirectly through more attracted advertisers. On the other hand, platforms want to increase the level of user data collected as it enhances targeting, attracts advertisers and hence increases profits. In equilibrium, stronger competition for users impacts the former effect of attracting users more than the latter of increasing targeting, therefore, platforms will collect less user data. Following the same intuition, platforms would be willing to lose some advertisers in order to not repel valuable users. Hence, also equilibrium advertiser prices increase in t_u (compare Online Appendix B.1). Contrary to the effects of advertiser-side competition, these results reflect the 'standard' two-sided platform logic: stronger competition for users reduces the 'price' on the user side, while it increases the price on the advertiser side.

Furthermore, we discuss the effect of user-side competition intensity on platform profits. One could expect that platforms' profit increases if competition for users becomes less intense, however the opposite is true. For this note that their profit function in equilibrium is $\Pi_i^* = p^* \tau(d^*)\, A_i^*\, X_i^* = (1/4)\, p^* \tau(d^*)$. Then, a change in user-side competition intensity t_u gives

$$\frac{\mathrm{d}\Pi_i^*}{\mathrm{d}t_u} = \frac{1}{4}\left[\frac{\mathrm{d}p^*}{\mathrm{d}t_u}\tau(d^*) + \tau'(d^*)\frac{\mathrm{d}d^*}{\mathrm{d}t_u}p^*\right]. \tag{16}$$

On the one hand, advertiser prices decrease if competition for users becomes less intense (t_u increases), which reduces platform profits. Hence the first term on the right-hand side of (16) is negative. On the other hand, the second term is positive, because when competition for users becomes less intense (t_u increases), more data can be collected from users, which leads to more effective ad targeting and therefore increased platform profits. As can be seen from the derivation in Appendix A, overall, the negative first-term effect is stronger in equilibrium, such that platforms suffer from weaker competition for users, i.e. $\mathrm{d}\Pi_i^*/\mathrm{d}t_u < 0$.

6 Policy implications

In this section we draw comparisons between the different outcomes outlined in Section 4 and present policy implications.

6.1 Comparison of outcomes

First, we want to compare the outcome of the efficiency benchmark with the market equilibrium outcome. If we compare the right-hand-side of the competitive level d^* in (12) and the efficient level d^o in (7) we can see that the difference will crucially depend on the distortion induced by

$$\delta(d^*) := \frac{\nu(d^*) + t_u}{\tau(d^*) - t_a}, \tag{17}$$

which gives more or less weight to the marginal benefit on the advertiser market side $\tau'(d^*)/2$. Note that by Assumption 1 the denominator of $\delta(d^*)$ is positive, so that we have $\delta(d^*) > 0$. As the efficient level d^o does not depend on parameter values, we can see that there can be underprovision ($d^*_u < d^o$) as well as overprovision ($d^*_o > d^o$) of personal data in the competitive equilibrium. Depending on the structure of the market too much or too little weight is put on the advertiser side of the market. In particular we can infer from equations (12) and (7) that the competitive outcome leads to underprovision of personal data if $\delta(d^*) < 1$ and to overprovision if $\delta(d^*) > 1$. Note for $\delta(d^*) = 1$ expression (12) simplifies to (7), the efficient level of data provision. Using our definition of $\delta(d^*)$ we can then see that $d^* < d^o$ if

$$\delta(d^*) < 1 \quad \Longleftrightarrow \quad \tau(d^*) - \nu(d^*) > t_a + t_u \tag{18}$$

and $d^* > d^o$ if

$$\delta(d^*) > 1 \quad \Longleftrightarrow \quad \tau(d^*) - \nu(d^*) < t_a + t_u. \tag{19}$$

These results are summarized in the following proposition.

Proposition 4 *The competitive outcome leads to overprovision of personal data if com-*

petition on both market sides is weak and/or if net cross-group externalities are small. If competition on both market sides is strong and/or net cross-group externalities are large, the competitive outcome exhibits underprovision of personal data.

Proof. See Appendix A.4. ■

We want to interpret this finding by first holding the functions $\kappa(d)$, $\nu(d)$ and $\tau(d)$ fixed and asking the question which competitive environment leads to which scenario. From our comparative statics results we know that the amount of data is a monotone function of the transportation cost parameters, i.e. $\frac{\mathrm{d}d^*}{\mathrm{d}t_u} > 0$ and $\frac{\mathrm{d}d^*}{\mathrm{d}t_a} > 0$. Proposition 4 then gives us a threshold for how the resulting level of data collection compares to the efficient benchmark: if competition is too strong, i.e. $t_a + t_u$ is small, platforms tend to collect and process an inefficiently small amount of data as users and advertisers shy away too easily. If in turn competition on both sides is weak, i.e. $t_a + t_u$ is high, the market sides become more sticky and platforms are able to extract high amounts of personal data.

We can also hold the competitive environment t_a, t_u on both sides fixed and analyze the effects of relatively strong or weak opposing cross-group externalities. On the one hand, an additional user imposes a positive externality on advertisers (and platforms), which is equal to the targeting effect $\tau(d^*)$. On the other hand, an additional advertiser imposes a negative externality on users, which is equal to the nuisance costs $-\nu(d^*)$. The net effect can therefore be interpreted as available gains from trade in this economy. If the net effect is relatively large, there are significant gains of trade which could be seized by increasing the amount of data collected. If the net effect is small, the gains from trade could be increased by lowering the amount of collected data.

Comparing the user-optimal level d^u to the welfare-optimal level d^o we immediately see that users would provide an inefficiently low level of data. This result is summarized in the following proposition.

Proposition 5 *The user-optimal level of data provision is inefficiently low.*

The reason for this result is straightforward. As users do not internalize the effect the data has on the advertiser market, they will provide data up to the point where the marginal decrease in nuisance equals marginal cost of data provision. Since from a welfare perspective the value creation aspect on the advertiser market is omitted, the resulting level of data provision is inefficiently low. Furthermore, since $\delta(d^*) > 0$ we also have $d^* > d^u$ for all exogenous parameters and functional forms, as shown in Corollary 1. Unlike users, platforms act as intermediaries and are able to internalize parts of the value creation on both sides of the market.

6.2 Policy conclusions

In this subsection we briefly discuss what conclusions can be drawn from our previous analyses when it comes to policy implications and regulation.

In our model, an omnipotent regulator could obviously achieve the first-best outcome by forcing $d_i = d_j = d^o$ and increasing competition on both sides of the market such that $t_u \to 0$ and $t_a \to 0$. In this case the efficient amount of data is provided while the total transportation costs approach zero.

In practice, regulation and policy discussions typically focus on data and privacy regulation or on competition policy measures (or merger regulation) to assure competitiveness on the user side, for example in the recent Facebook case at the BKartA or the Facebook/Whatsapp merger case in the US and the EU. In this section we want to present answers our model provides for privacy and competition policy, taking into account both market sides and at the same time the effect on privacy.

Privacy regulation

Holding the competitive structure of the market fixed, the regulator could improve upon the market outcome by enforcing the efficient level of private data provision $d_i = d_j = d^o$. However, a direct regulation of the amount of data in our model requires knowledge of the cross-group externalities, i.e. functions $\tau(d)$ and $\nu(d)$, as well as users' privacy concerns $\kappa(d)$.

A regulator could also consider switching to a consumer standard and let consumer freely choose how much data they would like to provide. Our results show that the user-optimal amount of data is always inefficiently low as users do not internalize the benefit on the advertiser side. In particular our results suggest that we can only improve in terms of welfare by switching to a consumer standard when there is extreme overprovision of data in the economy, i.e. platforms have significant market power on both sides of the market.

If the market exhibits underprovision, switching to the consumer standard always reduces welfare.

Competition policy

An approach which is less demanding when it comes to information requirements is the regulation of the competitive environment on both market sides, i.e. t_u and t_a. Our results (Proposition 4) suggest that if competition is very weak on both sides ($t_u + t_a$ high), the amount of data collected is likely to be inefficiently high. Similarly, if competition is too strong (t_u+t_a low), too little data is provided from a welfare point of view. While regulators still have to know whether there is overprovision or underprovision in the market in the first place, our results can still provide some guidance.

Our comparative statics results suggest that increasing competition works in the same direction for both sides of the market. The equilibrium amount of data provision is a monotone function of the transportation cost parameters t_a and t_u and by altering either one of the parameters it is possible to push the competitive equilibrium amount of data d^* towards the welfare optimum d^o. Typical examples include reducing switching costs on the user side (see e.g. GDPR/data portability in the EU) or policing vertical integration on the advertiser side (see e.g. debate around Google/DoubleClick acquisition). Further, our results suggest that more competition between platforms is not necessarily welfare enhancing as it further limits the ability to create economic value through the collection of personal data in the case of underprovision.

Also, our results suggest that policy measures, although they work in the same direction, are not equally effective across market sides, i.e. $\frac{\mathrm{d}d^*}{\mathrm{d}t_a} \neq \frac{\mathrm{d}d^*}{\mathrm{d}t_u}$. This might be particularly important in a scenario where the market exhibits underprovision and a regulator would have to reduce competition as this implies increasing transportation costs in the economy. Increasing transportation costs would then lead to more data collection in the subsequent market outcome. Whether we can increase total welfare by increasing transportation costs, however, depends crucially on whether the benefit of higher and thus more efficient data provision (non linear) exceeds the increased costs of transportation (linear).[20] This trade-

[20]Note that also in a situation of overprovision, the market structure might be such that it is socially beneficial to decrease transportation costs, i.e. increase competition, even beyond the level where it induces efficient data provision (as established in equation 7), such that the benefits of decreased transportation

off could call for a second-best regulation, where competition intensity is regulated in such a way that the amount of data provided in the subsequent market outcome balances the above mentioned benefits and costs at the margin.

From these results on competition policy we want to draw two main conclusions. First, regulating competition on either or both market sides can address the privacy / data collection distortion in the market outcome. Second, whenever regulators consider competition policy or merger regulation in these data-driven industries, they should take into account the impact on data collection in the market.

7 Discussion

In this chapter we sketch and briefly discuss extensions and variations of the baseline model presented in Section 3.

7.1 User prices

In this section we consider an alternative setup where platforms can charge prices on the user side of the market. All other model specifications remain as before, i.e. specifically users now have to pay a monetary price additional to their personal data 'payment'. In a sense, this setup could be considered as an unrestricted model, where platforms are not restricted to zero user prices. Let p_i^u denote the price a user has to pay to join platform i. User utility is then given by

$$u_i(x) = \underline{v}_i + \underline{d} - \kappa(d_i) - \nu(d_i)A_i - p_i^u - t_c|l_i - x|, \tag{20}$$

while advertisers still face the same decision as in Section 3. Market shares are obtained as before by pinning down indifferent users and advertisers and solving the resulting system of equations. The resulting profit maximization problem of platform i is then given by

$$\max_{p_i, d_i, p_i^c} = A_i\tau(d_i)p_iX_i + p_i^uX_i \ \forall i \in \{1, 2\}. \tag{21}$$

costs outweigh the costs from data underprovision.

Following the same procedure as in our baseline model we obtain symmetric equilibrium values $p_i = p_j = \tilde{p}$, $p_i^u = p_j^u = \tilde{p}^u$ and $d_i = d_j = \tilde{d}$ where advertiser prices are given by $\tilde{p} = 2[t_a + \nu(\tilde{d})]/\tau(\tilde{d})$, user prices by

$$\tilde{p}^u = t_a + t_c + \nu(\tilde{d}) - \tau(\tilde{d}), \tag{22}$$

while the equilibrium amount of data is given by

$$\kappa'(\tilde{d}) = \frac{1}{2}\left[\tau'(\tilde{d}) - \nu'(\tilde{d})\right]. \tag{23}$$

We immediately see from equations (7) and (23) that $\tilde{d} = d^o$.

Proposition 6 *If platforms can charge prices on both market sides, the efficient level of data is collected.*

Since platforms can now extract rents from both sides of the market, they maximize the aggregate value, whereas in our baseline model platforms only profited on the advertiser side of the market and hence set a data requirement level which is distorted. Taking a closer look at equilibrium user prices in (22) we immediately see that negative, positive or zero user prices are possible, depending on parameter values and functional forms.

Proposition 7 *If user prices in the two-sided pricing model are positive, the one-sided pricing constraint would result in data overprovision. Contrary, if user prices are negative, this constraint would yield underprovision.*

Proof. See Appendix A.4. ■

The intuition for this result is that now platforms can extract the efficient amount of data by adequately compensating users. If net benefits of data collection are large or competition is rather strong, platforms can extract large amounts of data from users and then compensate them by charging negative user prices, whereas in the one-sided pricing model platforms do not have the instrument for compensation and therefore are forced to collect less data than the efficient level. Vice versa, if net benefits are small or competition rather weak, platforms are not forced to monetize through ads by extracting an inefficiently high amount of data, but can obtain positive revenue from the user side instead and leave the amount of data at the efficient level.

We would like to mention at this point that this result may depend on the fact that even with positive user prices we assume the user market to remain fully covered. However, remember that under a market solution with overprovision users gain in terms of utility by decreasing d from d^* to d^o. If this difference in utility is enough to cover the associated positive user price, the user market remains covered. If the consumer price exceeds the utility gain, the two-sided pricing may lead to users leaving the market and efficiency may not be feasible any longer. We provide a more detailed discussion of the full market coverage assumption in the subsequent section. A similar argument can also be made if we consider heterogeneous users as then our uniform pricing setup may not be sufficient to ensure efficiency but platforms would need to engage in price discrimination.

Nevertheless, we would like to draw two further conclusions from these results. Firstly, observing a user price $\tilde{p}^u = 0$ empirically is consistent with the equilibrium result above as well as with our baseline model. By observing zero prices we can not infer whether a price of zero is an optimal choice, making the model above the 'correct' model, or whether there are constraints which prevent platforms from setting user prices at all, making our baseline model more suitable. Secondly, since user prices depend on parameters of competition intensity and externalities, observing zero prices across different markets, jurisdictions and industry sectors makes it unlikely that $\tilde{p}^u = 0$ is a profit maximizing choice in all cases. This strongly supports the argument made by Waehrer (2015) that user prices are not a (practical) variable of interest in real-world platform maximization problems.

7.2 Collusion

Full collusion

Let us consider a collusive game where platforms agree on prices $p_i = p_j = p$ and data requirements $d_i = d_j = d$ such that joint profits are maximized. Since advertisers face transportation costs, the profit maximizing collusive price is such that the participation constraint of the indifferent advertiser is binding $\pi_i\left(\frac{1}{2}\right) = 0$ which yields $p = 1 - \frac{t_a}{\tau(d)}$. Plugging the collusive price p into the platforms' profit functions (3) we obtain $\Pi_i =$

$\frac{1}{4}(\tau(d) - t_a)$ and immediately see that profits are increasing in d up to the point where the participation constraint of the indifferent user binds $d : u_i(\frac{1}{2}) = 0$. Since we assumed $\underline{u}$ to be high enough to have interior solutions in the previous sections, we can infer that the collusive amount of data will be excessively high.

Partial collusion

In this section we consider an alternative collusive environment where platforms coordinate on setting a symmetric level of data d but still compete in prices on the advertiser market. The idea is that platforms might influence privacy regulation in a collusive effort without coordinating their pricing decisions. We therefore introduce a collusive stage where platforms agree on a symmetric level d prior to the price setting decision. It is easy to verify that symmetric prices are then given by $p_i = p_j = p(d) \equiv 2\frac{t_a t_u + \nu(d)\tau(d)}{\tau(d)[t_u + \nu(d)]}$, similar to the market outcome outlined in Section 4. The key difference, however, is the collusive choice of d. As prices (and d) are symmetric, market shares can be anticipated to be given by $A_i = A_j = X_i = X_j = 1/2$ such that industry wide platform profits are given by $\Pi(d) := \Pi_i(d) + \Pi_j(d) = \frac{\tau(d)p(d)}{2}$.

If we have $\Pi'(d) > 0$ for all d, the collusive level will be the same as in full collusion case, such that the participation constraint of the users will be binding, and if $\Pi'(d) = 0$ has a solution, a possible interior solution exists. The comparison to the market outcome (or to the efficient outcome) is in this case, however, ambiguous and depends on functional forms and parameter values.

Interestingly, industry profits are not necessarily increasing in d. In fact if $\Pi'(d) = p(d)\tau'(d) + p'(d)\tau(d) < 0$ for all d then the collusive level of data will be zero. The reason for this seemingly counter-intuitive result is that increasing d can effectively propagate competition on the advertiser market. In particular if we go back to the definition of advertiser market shares in (5) we can see that increasing a symmetric level d has the same effect on the advertiser market as a decrease in transportation costs in the sense that it makes advertisers more reactive towards changes in prices. The intuition is straightforward: if the click-probability is very high, small differences in prices become magnified. The trade-off faced by the platforms is then the following. An increase in click probability (through increasing d) results in tighter competition on the ad market (depressing p). The optimal d can therefore vary widely depending on which effect dominates.

To briefly summarize this section we can conclude that full collusion amongst platforms should be avoided whenever possible. When it comes to partial collusion, however, a more nuanced analysis is necessary as competition on the ad market might be sufficiently strong to prevent inefficient regulatory capture.

7.3 Market coverage and multi-homing

In this section we want to briefly discuss the effects of relaxing the assumptions guaranteeing full market coverage and single-homing. We consider market-coverage and multi-homing together because without these assumptions in both cases the market share of a platform is determined by the user/advertiser who is indifferent between joining a platform and the outside option, whereas in the baseline model it was determined by the user/advertiser who is indifferent between joining both platforms. Note that this changes the interpretation of transportation costs in the model substantially. While in the baseline model transportation costs measured a restriction to switching to the other platform and hence a degree of platform competition, now they rather exhibit a restraint on a platform's demand, independent of the other platform. Essentially, lower transportation costs can now be interpreted as *more* elastic demand, whereas in the baseline model they reflected *less* elastic (sticky) demand. While our assumptions for the baseline model were chosen to study full competition between platforms, relaxing the assumptions on one market side significantly changes the setting in the sense that platforms now only compete indirectly through the other market side. Nevertheless, we want to provide some intuition for the robustness of our results. For a more detailed analysis consider the Online Appendix B.

Advertiser side

On the advertiser side, lifting Assumption 1 of a covered market together with the single-homing assumption can result in two cases, depending on parameters. First, if transportation costs t_a are sufficiently small, some advertisers 'in the middle' will use both platform (multi-homing). The comparison of the new equilibrium level of data provision to the new efficient level or the baseline level of data provision is, however, ambiguous. This is because less advertiser demand elasticity on the one hand could allow firms to readjust d, while at the same time the total number of advertisers on a platform could rise. From an efficiency perspective, though, more data should be collected than was efficient in the baseline model. Second, if transportation costs t_a are sufficiently high, some advertisers in the middle might choose not to use any platform (no full market coverage). Then it would also be efficient to exclude some advertisers such that the new efficient level of data provision is below the efficient baseline level. The comparison to the equilibrium outcomes remains however ambiguous, as above.

User side

On the user side, relaxing the full-market Assumption 2 and the single-homing constraint similarly leads to either some users 'in the middle' joining both platforms (multi-homing) or some user joining neither platform (no full market coverage), depending mainly on transportation costs t_u. In both cases user demand is then merely scaled by the demand elasticity, i.e. the transportation costs t_u, and users' role essentially reduces to being a resource of data needed to create advertising surplus.[21] We find that there would always be over-provision of user data in equilibrium because the efficient benchmark takes into account the trade-off between total value creation and user exclusion, whereas the market outcome only balances targeting benefits and potential user exclusion. However, still less data is collected than in the baseline model and also the efficient level of data decreases. Further, we find that now the transportation costs parameters have no effect on the equilibrium level of data provision. This is because t_u merely scales demand while the relevant trade-off for the choice of d involves the actual utility when joining the platform and is not influenced by the demand scale. Furthermore, equilibrium prices now increase in t_u and decrease in t_a. Because of the reversed role transportation costs now play, this is not contradictory to the baseline model results: the harder it is to keep users, the higher the price for advertisers. Consequently, platform profits still increase and advertiser profits still decrease in user-side elasticity.

[21] Note that on the advertiser side this was not the case because advertisers pay money rather than a value-creating resource.

7.4 Positive cross-group externalities

In the baseline model we considered the case where users incur nuisance cost from seeing ads on the platform, i.e. a negative cross-group externality incurred by users. As explained in the beginning we consider this case because we think it illustrates the main results in a very intuitive way. What we demonstrate in the Online Appendix B is that the model can in fact be generalized to have positive cross-group effects in both directions while the major results remain unchanged.

8 Conclusion

We analyze the role of competition intensity in a two-sided market framework where platforms collect data from users and monetize through ad-sales. Our model predicts that the equilibrium amount of collected data will be distorted compared to the welfare efficient benchmark. Depending on the net effect of cross-group externalities and the competition intensity on both sides of the market, the distortion can lead to underprovision or overprovision of personal data. Since the level of collected data increases the more market power platforms have on either side of the market, side specific regulations are substitutable. We also show that a consumer standard would always lead to underprovision of data as users do not internalize improvements in the targeting capabilities. Lastly, we showed that two-sided pricing induces platforms to choose the efficient level of data by adequately compensating users.

While we think our model provides useful insights we would also like to discuss some limitations. It would be interesting to further explore the role of multi-homing on the advertiser side as it changes the competitive dynamics substantially. Secondly, one could alter the setting on the user side and consider heterogeneous users, while platforms engage in second degree discrimination by offering a menu of data choices. We think those are interesting avenues for future research.

Appendix

A Omitted proofs

A.1 Second stage market shares

Note that equations (4) - (5) are consistent, non-redundant and linear in X_i, X_j, A_i, A_j such that the resulting solution in (4.1) is unique. Explicit market shares are then given by:

$$X_i = \frac{t_a(2\kappa(d_j) - 2\kappa(d_i) + \nu(d_j) - \nu(d_i) + 2t_u) + (1-p_j)\tau(d_j)(\nu(d_i) + \nu(d_j))}{4t_a t_u + (\nu(d_i) + \nu(d_j))((1-p_i)\tau(d_i) + (1-p_j)\tau(d_j))}$$
$$X_j = \frac{t_a(2\kappa(d_i) - 2\kappa(d_j) + \nu(d_i) - \nu(d_j) + 2t_u) + (1-p_i)\tau(d_i)(\nu(d_i) + \nu(d_j))}{4t_a t_u + (\nu(d_i) + \nu(d_j))((1-p_i)\tau(d_i) + (1-p_j)\tau(d_j))}$$
$$A_i = \frac{(1-p_i)\tau(d_i)(\kappa(d_j) - \kappa(d_i) + \nu(d_j) + t_u) - (1-p_j)\tau(d_j)(\kappa(d_i) - \kappa(d_j) - \nu(d_j) + t_u) + 2t_a t_u}{4t_a t_u + (\nu(d_i) + \nu(d_j))((1-p_i)\tau(d_i) + (1-p_j)\tau(d_j))}$$
$$A_j = 1 - \frac{(1-p_i)\tau(d_i)(\kappa(d_j) - \kappa(d_i) + \nu(d_j) + t_u) - (1-p_j)\tau(d_j)(\kappa(d_i) - \kappa(d_j) - \nu(d_j) + t_u) + 2t_a t_u}{4t_a t_u + (\nu(d_i) + \nu(d_j))((1-p_i)\tau(d_i) + (1-p_j)\tau(d_j))}$$

A.2 Second order conditions

In the following we derive sufficient conditions such that the equilibrium values $p^*,\ d^*$ derived from the maximization problem presented in Section 3 characterize a local maximum. Let us consider the Hessian evaluated at equilibrium values. Starting with

$$\left.\frac{\partial^2 \Pi_i}{\partial p_i^2}\right|_{d^*,p^*} = -\frac{t_u^2\ \tau(d^*)^2(\nu(d^*) + t_u)}{4(t_u - \nu(d^*))^2(\nu(d^*)\tau(d^*) + t_a t_u)}$$

we immediately see that $\left.\frac{\partial^2 \Pi_i}{\partial p_i^2}\right|_{d^*,p^*} < 0$, a necessary condition for the Hessian to be negative definite. In the next steps we argue that we can always find functions $\tau(\cdot), \nu(\cdot)$ such that $det(H)|_{d^*,p^*} > 0$. First, it is helpful to look at the numerator and the denominator of the Hessian separately

$$det(H)|_{d^*,p^*} = \frac{H_{num}}{H_{den}}$$

where the numerator H_{num} and the denominator H_{den} are given by

$$\begin{aligned} H_{num} = \tau(d^*)^2 \big[& -4t_u^2(t_a - \tau(d^*))(\nu(d^*)\tau(d^*) + t_a t_u)\left(\nu''(d^*)(t_a - \tau(d^*)) + \tau''(d^*)(\nu(d^*) + t_u)\right) \\ & - t_u^2 \nu'(d^*)^2 (t_a - \tau(d^*))^3 - \tau'(d^*)^2(\nu(d^*) + t_u)^2\left(\nu(d^*)(\nu(d^*)(t_a - \tau(d^*)) + 4t_c\tau(d^*)) + 4t_a t_u^2\right) \\ & + 2t_u\nu(d^*)\nu'(d^*)\tau'(d^*)(t_a - \tau(d^*))^2(\nu(d^*) + t_u)\big] \\ H_{den} = {} & 64(t_a - \tau(d^*))(t_u - \nu(d^*))^2(\nu(d^*)\tau(d^*) + t_a t_u)^2 \end{aligned}$$

Note that $H_{den} < 0$ as we have $(t_a - \tau(d^*)) < 0$ from Assumption 1. Rewriting H_{num} as

$$
\begin{aligned}
H_{num} &= \tau(d^*)^2 \left[H1_{num}\left(H2_{num}\nu''(d^*) + H3_{num}\tau''(d^*)\right) + H4_{num} + H5_{num} + H6_{num}\right] \\
H1_{num} &= -4t_u^2(t_a - \tau(d^*))(\nu(d^*)\tau(d^*) + t_a t_u) > 0 \\
H2_{num} &= (t_a - \tau(d^*)) < 0 \\
H3_{num} &= (\nu(d^*) + t_u) > 0 \\
H4_{num} &= -t_u^2\nu'(d^*)^2(t_a - \tau(d^*))^3 \geq 0 \\
H5_{num} &= -\tau'(d^*)^2(\nu(d^*) + t_u)^2\left(\nu(d^*)(\nu(d^*)(t_a - \tau(d^*)) + 4t_u\tau(d^*)) + 4t_a t_u^2\right) \lesseqgtr 0 \\
H6_{num} &= 2t_u\nu(d^*)\nu'(d^*)\tau'(d^*)(t_a - \tau(d^*))^2(\nu(d^*) + t_u) \leq 0
\end{aligned}
$$

we can see that requiring $H_{num} < 0$ is equivalent to the condition

$$
-\frac{1}{H1_{num}}\left(H4_{num} + H5_{num} + H6_{num}\right) > H2_{num}\nu''(d^*) + H3_{num}\tau''(d^*)
$$

where $LHS \lesseqgtr 0$ while $RHS < 0$ due to our functional requirements on $\tau(\cdot)$ and $\nu(\cdot)$. The important thing to realize is that, firstly, the condition for negative definiteness reduces to a condition which is linear in $\nu''(d^*)$ and $\tau''(d^*)$, the curvature information of the targeting and the nuisance functions, and secondly, is given by an upper bound. If the sign of the upper bound is positive then this condition is always fulfilled as we have $RHS < 0$. Only if the sign of the upper bound is negative, the condition may bind. But then we can assume that $\tau(\cdot)$ is sufficiently concave and/or $\nu(\cdot)$ is sufficiently convex such that this condition holds since for our results we only require $\tau''(\cdot) < 0$ and $\nu''(\cdot) \geq 0$ which is in line with this condition.

A.3 Market outcome

In equilibrium $p^* < 1$ and $\pi_i^*(a) \geq 0$. To see this note that given equation (13), $p^* < 1$ if

$$
2\frac{t_a t_u + \nu(d^*)\tau(d^*)}{\tau(d^*)t_u + \nu(d^*)\tau(d^*)} < 1 \iff t_a < \tau(d^*)\frac{(t_u - \nu(d^*))}{2t_u} < \tau(d^*) \tag{A.1}
$$

By Assumption 1 we have that $\tau(d) > t_a$ for all d and therefore in particular also $\tau(d^*) > t_a$. Further, we have that $0 < (t_u - \nu(d^*))/2t_u < 1$, hence the last inequality. Thus, Assumption 1 is sufficient for the expression above to hold and $p^* < 1$.

Even the indifferent advertiser with highest transportation costs has positive profits in equilibrium because

$$\pi_i^*(\frac{1}{2}) = \frac{\tau(d^*)}{2} - \frac{t_a t_u + \nu(d^*)\tau(d^*)}{t_u + \nu(d^*)} - \frac{t_a}{2} \geq 0 \iff \tau(d^*)\frac{t_u - \nu(d^*)}{3t_u + \nu(d^*)} \geq t_a, \tag{A.2}$$

which is guaranteed by Assumption 1 for all d and especially for d^*. For this note that the term on the left in the last inequality is increasing in d.

A.4 Proofs of Propositions & Corollaries

Proof of Corollary 1

Proof. Rearranging terms in the first-order condition of platform profit maximization, given by equation (12), yields $2\kappa'(d^*) + \nu'(d^*) = \tau'(d^*)\frac{\nu(d^*)+t_u}{\tau(d^*)-t_a}$. By Assumption 1 we have $\tau(d^*) > t_a$. Hence the right hand side is positive, such that $2\kappa'(d^*) + \nu'(d^*) > 0$. ■

Proof of Proposition 4

Proof. The proof relies on the monotonicity of the LHS and RHS in equations (7) and (12). Suppose, $\delta(d^*) > 1$ but $d^* < d^o$ and hence $\kappa'(d^*) < \kappa'(d^o)$. Using the implicit definition of d^o in (7) and d^* in (12) this implies $\delta(d^*)\tau'(d^*) - \nu'(d^*) < \tau'(d^o) - \nu'(d^o)$. Rearranging yields $\delta(d^*) < \frac{\tau'(d^o)}{\tau'(d^*)} + \frac{\nu'(d^*)-\nu'(d^o)}{\tau'(d^*)}$. But due to the curvature of $\tau(\cdot), \nu(\cdot)$ we have $\frac{\tau'(d^o)}{\tau'(d^*)} < 1$ and $\frac{\nu'(d^*)-\nu'(d^o)}{\tau'(d^*)} \leq 0$ for $d^* < d^o$, contradicting $\delta(d^*) > 1$. Now suppose $\delta(d^*) > 1$ and $d^* > d^o$, and hence $\delta(d^*) < \frac{\tau'(d^o)}{\tau'(d^*)} + \frac{\nu'(d^*)-\nu'(d^o)}{\tau'(d^*)}$. For $d^* > d^o$ we then have $\frac{\tau'(d^o)}{\tau'(d^*)} > 1$ and $\frac{\nu'(d^*)-\nu'(d^o)}{\tau'(d^*)} \geq 0$ and hence $\delta(d^*) > 1$. ■

Proof of Proposition 7

Proof. To see that positive user prices in the two-sided model correspond to data overprovision in the one-sided pricing model, note that user prices are positive in the two-sided pricing model if $\tau(d^o) - \nu(d^o) < t_a + t_u$. From Proposition 4 we know that in the one-sided pricing model too little data is provided if $t_a + t_u < \tau(d^*) - \nu(d^*)$. But this would mean that $d^* < d^o$, which contradicts $\tau(d^o) - \nu(d^o) < t_a + t_u < \tau(d^*) - \nu(d^*)$, as $\tau(d)$ is increasing and $\nu(d)$ decreasing in d. Hence it can only be that in the one-sided model there is overprovision, such that $d^* > d^o$ and $\tau(d^o) - \nu(d^o) < \tau(d^*) - \nu(d^*) < t_a + t_u$.

To see that negative user prices in the two-sided model correspond to data underprovision in the one-sided pricing model, note that user prices are negative in the two-sided pricing model if $\tau(d^o) - \nu(d^o) > t_a + t_u$. From Proposition 4 we know that too much data is provided if $t_a + t_u > \tau(d^*) - \nu(d^*)$. But this would mean that $d^* > d^o$, which contradicts $\tau(d^o) - \nu(d^o) > t_a + t_u > \tau(d^*) - \nu(d^*)$. Hence it must be that in the one-sided model there is underprovision, such that $d^* < d^o$ and $\tau(d^o) - \nu(d^o) > \tau(d^*) - \nu(d^*) > t_a + t_u$. ■

A.5 Proofs for comparative statics

For the effect of transportation costs t_a and t_u on d_i note first that

$$\left.\frac{\partial X_i}{\partial d_i}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = -\frac{\{t_a t_u + \nu(d^*)\left[t_a p^* + (1-p^*)\tau(d^*)\right]\}\,\tau'(d^*)}{4\left[\tau(d^*) - t_a\right]\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} < 0, \tag{A.3}$$

$$\left.\frac{\partial A_i}{\partial d_i}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = -\frac{(1-p^*)\left[t_a t_u + \nu(d^*)\tau(d^*)\right]\tau'(d^*)}{4\left[\tau(d^*) - t_a\right]\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} < 0, \tag{A.4}$$

because $\tau'(d^*) > 0$, while $\tau(d^*) > t_a$ by Assumption 1 and $p^* < 1$ as established in Section A.3. Differentiating (A.3) and (A.4) with respect to transportation costs yields

$$\left.\frac{\partial^2 X_i}{\partial d_i \partial t_a}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = -\frac{(1-p^*)\nu(d^*)\left[t_a t_u + \nu(d^*)\tau(d^*)\right]\tau'(d^*)}{4\left[\tau(d^*) - t_a\right]\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]^2} < 0,$$

$$\left.\frac{\partial^2 A_i}{\partial d_i \partial t_a}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = \frac{(1-p^*)t_u\left[t_a t_u + \nu(d^*)\tau(d^*)\right]\tau'(d^*)}{4\left[\tau(d^*) - t_a\right]\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} > 0,$$

$$\left.\frac{\partial^2 X_i}{\partial d_i \partial t_u}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = \frac{t_a\{t_a t_u + \nu(d^*)\left[t_a p^* + (1-p^*)\tau(d^*)\right]\}\,\tau'(d^*)}{4\left[\tau(d^*) - t_a\right]\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} > 0,$$

$$\left.\frac{\partial^2 A_i}{\partial d_i \partial t_u}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = \frac{(1-p^*)\tau(d^*)\{t_a t_u + \nu(d^*)\left[t_a p^* + (1-p^*)\tau(d^*)\right]\}\,\tau'(d^*)}{4\left[\tau(d^*) - t_a\right]\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} > 0.$$

Further note that

$$\left.\frac{\partial^2 A_i}{\partial d_i \partial t_a} - \frac{\partial^2 X_i}{\partial d_i \partial t_a}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = \frac{(1-p^*)\left[t_u + \nu(d^*)\right]\left[t_a t_u + \nu(d^*)\right]\tau'(d^*)}{4\left[\tau(d^*) - t_a\right]\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} > 0.$$

To see that $\mathrm{d}\Pi_i^P/\mathrm{d}t_u < 0$, note that

$$\begin{aligned}\frac{\mathrm{d}\Pi_i^P}{\mathrm{d}t_u} &= \frac{-\left[\tau(d^*) - t_a\right]\left[\nu(d^*) - t_u\nu'(d^*)\frac{\mathrm{d}d^*}{\mathrm{d}t_u}\right] + \frac{\mathrm{d}d^*}{\mathrm{d}t_u}\nu(d^*)\tau'(d^*)\left[t_u + \nu(d^*)\right]}{\left[t_c + \nu(d^*)\right]^2} \\ &= \frac{\left[\tau(d^*) - t_a\right]^2\left[(t_u + \nu(d^*))\,\nu'(d^*)\tau'(d^*) - \nu(d^*)\left\{(\tau(d^*) - t_a)\left[\kappa''(d) + \nu''(d^*)\right] - t_c\left(t_u + \nu(d^*)\right)\tau''(d^*)\right\}\right]}{\left[t_u + \nu(d^*)\right]^2\Psi(d^*)} \\ &< 0,\end{aligned} \tag{A.5}$$

where $\mathrm{d}d^*/\mathrm{d}t_u$ is from equation (15), while the term in the denominator is given by

$$\Psi(d^*) = \left[2\kappa''(d^*) + \nu''(d^*)\right] (\tau(d^*) - t_a)^2 - \nu'(d^*)\tau'(d^*)\,(\tau(d^*) - t_a) \\ + (\nu(d^*) + t_u)\left[\tau'(d^*)^2 - (\tau(d^*) - t_a)\,\tau''(d^*)\right] > 0. \tag{A.6}$$

Note for the inequalities that $\tau'(d^*) > 0$, $\tau''(d^*) < 0$ while $\nu'(d^*) \leq 0$, $\nu''(d^*) \geq 0$ by construction, and $\tau(d^*) > t_a$ by Assumption 1.

References

Anderson, S. P. and Coate, S. (2005). Market provision of broadcasting: A welfare analysis. *The review of Economic studies*, 72(4):947–972.

Anderson, S. P. and Gabszewicz, J. J. (2006). The media and advertising: a tale of two-sided markets. *Handbook of the Economics of Art and Culture*, 1:567–614.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics*, 37(3):668–691.

Armstrong, M. and Wright, J. (2007). Two-sided markets, competitive bottlenecks and exclusive contracts. *Economic Theory*, 32(2):353–380.

Bloch, F. and Demange, G. (2017). Taxation and privacy protection on internet platforms. *Journal of Public Economic Theory*.

Bonneau, J. and Preibusch, S. (2010). The privacy jungle: On the market for data protection in social networks. *Economics of information security and privacy*, pages 121–167.

Bourreau, M., Caillaud, B., and De Nijs, R. (2017). Taxation of a digital monopoly platform. *Journal of Public Economic Theory*. forthcoming.

Casadesus-Masanell, R. and Hervas-Drane, A. (2015). Competing with privacy. *Management Science*, 61(1):229–246.

Cooper, J. C. (2013). Privacy and antitrust: Underpants gnomes, the first amendment, and subjectivity. *George Mason Law & Economics Research Paper*, (13-39).

De Corniere, A. and De Nijs, R. (2016). Online advertising and privacy. *The RAND Journal of Economics*, 47(1):48–72.

EDPS (2014). Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the digital economy. https://edps.europa.eu/sites/edp/files/publication/14-03-26_competitition_law_big_data_en.pdf. [Online; accessed 17-May-2017].

European Commission (2015). Why we need a digital single market. https://ec.europa.eu/commission/sites/beta-political/files/dsm-factsheet_en.pdf. [Online; accessed 17-May-2017].

Kummer, M. and Schulte, P. (2016). When private information settles the bill.

Laudon, K. C. (1996). Markets and privacy. *Communications of the ACM*, 39(9):92–104.

Lefouili, Y. and Toh, Y. L. (2017). Privacy and quality. *TSE Working Paper*, 17(795).

Peitz, M. and Reisinger, M. (2016). The economics of internet media. In *Handbook of Media Economics, vol. 1A*. Elsevier.

Peitz, M. and Valletti, T. M. (2008). Content and advertising in the media: Pay-tv versus free-to-air. *international Journal of industrial organization*, 26(4):949–965.

Pew Research Center (2014). Public perceptions of privacy and security in the post-snowden era. http://www.pewinternet.org/files/2014/11/PI_PublicPerceptionsofPrivacy_111214.pdf. [Online; accessed 17-May-2017].

Posner, R. A. (1981). The economics of privacy. *The American economic review*, 71(2):405–409.

Reisinger, M. (2012). Platform competition for advertisers and users in media markets. *International Journal of Industrial Organization*, 30(2):243–252.

Spiegel, Y. (2013). Commercial software, adware, and consumer privacy. *International Journal of Industrial Organization*, 31(6):702–713.

Stucke, M. E. and Grunes, A. P. (2016). *Big Data and Competition Policy*. Oxford University Press.

Waehrer, K. (2015). Online services and the analysis of competitive merger effects in privacy protections and other quality dimensions. *Available at SSRN*.

B Online appendix

In this online appendix we will provide derivations and intuition for comparative statics not covered in the main text. Further, we present extensions to our baseline model where we consider multi-homing, elastic total demand and positive cross-group externalities.

B.1 Comparative static effects on prices in equilibrium

For this analysis consider the platform's first-order condition in equation (11) and note that the price depends indirectly on the effects of p_i on advertiser and user market shares A_i and X_i as given in Section 4.1.

$$\left.\frac{\partial X_i}{\partial p_i}\right|_{\substack{d_i=d^*\\p_i=p^*}} = \frac{\nu(d^*)\tau(d^*)}{4\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} > 0, \tag{B.1}$$

$$\left.\frac{\partial A_i}{\partial p_i}\right|_{\substack{d_i=d^*\\p_i=p^*}} = -\frac{t_u\tau(d^*)}{4\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]} < 0, \tag{B.2}$$

because $p^* < 1$ as established in Appendix A.

Competition for advertisers

Note that in Section 5 we discussed that lower advertiser-side competition intensity increases the level of data collection in equilibrium, i.e. $\mathrm{d}d^*/\mathrm{d}t_a > 0$. Here we analyze the effects of competition intensity for advertisers on p^*. Differentiating (B.1) with respect to transportation costs t_a yields

$$\left.\frac{\partial^2 X_i}{\partial p_i \partial t_a}\right|_{\substack{d_i=d^*\\p_i=p^*}} = -\frac{t_u\tau(d^*)\nu(d^*)}{4\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]^2} < 0,$$

$$\left.\frac{\partial^2 A_i}{\partial p_i \partial t_a}\right|_{\substack{d_i=d^*\\p_i=p^*}} = \frac{t_u^2\,\tau(d^*)}{4\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]^2} > 0.$$

Further note that

$$\left.\frac{\partial^2 A_i}{\partial p_i \partial t_a} - \frac{\partial^2 X_i}{\partial p_i \partial t_a}\right|_{\substack{d_i=d^*\\p_i=p^*}} = \frac{t_u\left[t_u + \nu(d^*)\right]\tau(d^*)}{4\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]^2} > 0.$$

If competition for advertisers softens, i.e. transportation costs t_a increase, advertisers become less sensitive to changes in prices such that $\partial^2 A_i/(\partial p_i \partial t_a) > 0$. Consequently, users become more sensitive to prices (which repel advertisers) such that $\partial^2 X_i/(\partial p_i \partial t_a) < 0$. Overall, the former effect dominates the latter effect in magnitude. Consequently, and as $X_i^* = A_i^* = 1/2$, the right-hand-side of equation (11) increases in t_a such that the equilibrium price must rise, i.e.

$$\frac{\mathrm{d}p^*}{\mathrm{d}t_a} > 0. \tag{B.3}$$

Intuitively, higher advertiser transportation costs mean more sticky advertisers and hence decreased platform competition for advertisers. Therefore, it is straightforward that advertiser prices rise, which is line with standard intuition.

Competition for users

In Section 5 we discussed that lower competition intensity for users decreases platforms' equilibrium level of data collection, i.e. $\mathrm{d}d^*/\mathrm{d}t_u > 0$. Here we analyze the effects of competition intensity for users on p^*. Differentiating (B.1) with respect to transportation costs t_u yields

$$\left.\frac{\partial^2 X_i}{\partial p_i \partial t_u}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = -\frac{t_a \tau(d^*)\nu(d^*)}{4\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]^2} < 0,$$

$$\left.\frac{\partial^2 A_i}{\partial p_i \partial t_u}\right|_{\substack{d_i=d^*\\ p_i=p^*}} = -\frac{(1-p^*)\nu(d^*)\tau(d^*)^2}{4\left[t_a t_u + (1-p^*)\nu(d^*)\tau(d^*)\right]^2} < 0.$$

If competition for users softens, i.e. transportation costs t_u increase, users become less sensitive to changes in prices such that $\partial^2 X_i/(\partial p_i \partial t_u) < 0$. Consequently, advertisers, too, become less sensitive to prices (which now repel less users) such that $\partial^2 X_i/(\partial d_i \partial t_u) < 0$. Therefore the right-hand-side of equation (11) decreases in t_u such that the equilibrium price must fall, i.e.

$$\frac{\mathrm{d}p^*}{\mathrm{d}t_u} < 0. \tag{B.4}$$

Again, this is in line with standard platform intuition: advertiser prices fall if the user side becomes less sensitive (elastic).

Nuisance

First, we consider the effects of nuisance on data collection.[22] Totally differentiating the first-order conditions from equations (12) and (13) w.r.t. $\nu(d)$ and solving for $\mathrm{d}d^*/\mathrm{d}\nu(d)$ yields

$$\left.\frac{\mathrm{d}d^*}{\mathrm{d}\nu(d)}\right|_{d=d^*} = \frac{\left(\tau(d^*) - t_a\right)\tau'(d^*)}{\Psi(d^*)} > 0. \tag{B.5}$$

Second, we evaluate the effects of nuisance on p^*. Solving for $\mathrm{d}p^*/\mathrm{d}\nu(d)$ and dropping the argument d^* of $\nu(d^*)$ and $\tau(d^*)$ to abbreviate, yields

$$\left.\frac{\mathrm{d}p^*}{\mathrm{d}\nu(d)}\right|_{d=d^*} = \frac{-2t_u\left(t_a - \tau\right)^2\left[\tau\left(\tau''\left(\nu + t_u\right) - \left(\tau - t_a\right)\left(2\kappa'' + \nu''\right)\right) - \left(\nu + t_u\right)\tau'^2\right]}{\left(\nu + t_u\right)\tau^2\,\Psi(d^*)} > 0, \tag{B.6}$$

where $\Psi(d^*)$ is defined in equation (A.6). Intuitively, higher (absolute) nuisance results in lower user demand. To counterbalance this effect, platforms would increase ad prices as ads become relatively less attractive. Additionally, more user data would be collected in order to soften the nuisance increase. Interpreted from the point of view of users, they are now willing to incur marginally more privacy costs in order to obtain some nuisance reduction.

[22]Note that nuisance is a function in our model. To assess an increase in nuisance we treat it as fixed and consider an upward shift, without changing any curvature. For this we slightly abuse notation to stay consistent with the rest of our comparative statics, such that e.g. by $\mathrm{d}d^*/\mathrm{d}\nu(d)|_{d=d^*}$ we intuitively consider the effect of adding a positive constant c to the function, i.e. $\nu(d) + c$ where $c > 0$, on d^*.

Targeting

First, we consider the effects of the targeting technology on data collection.[23] Solving for $\mathrm{d}d^*/\mathrm{d}\tau(d)$ yields

$$\left.\frac{\mathrm{d}d^*}{\mathrm{d}\tau(d)}\right|_{d=d^*} = -\frac{(\nu(d^*) + t_u)\,\tau'(d^*)}{\Psi(d^*)} < 0. \tag{B.7}$$

Second, we evaluate the effects of nuisance on p^*. Solving for $\mathrm{d}p^*/\mathrm{d}\tau(d)$ and dropping again the argument d^* to abbreviate, yields

$$\left.\frac{\mathrm{d}p^*}{\mathrm{d}\tau(d)}\right|_{d=d^*} = \frac{2t_u\,(\tau - t_a)\,[\tau''\,(\nu + t_u)\,t_a - (\tau - t_a)\,[\nu'\tau' + t_a\,(2\kappa'' + \nu'')]]}{(\nu + t_u)\,\tau^2\Psi(d^*)} \gtrless 0. \tag{B.8}$$

platforms to create the same ad value with less personal data, hence in equilibrium platforms will compete to 'relax' the data requirement for users. Two effects are relevant for the effect on ad prices. On the one hand ads become more valuable, hence platforms might increase the price, i.e. their share, of this value (intensive margin). On the other hand, platforms might prefer to attract more of these valuable advertisers by reducing the ad price (extensive margin). Overall, the effect on ad prices depends on which of the opposing effects is stronger.

B.2 Comparative static effects on platform profits, advertiser profits and user utility

In this subsection we provide further intuition on equilibrium profits and utility by presenting comparative statics.

Effects on platform profits

The effects on platform profits $\Pi_i^* = p^*\tau(d^*)\,X_i^*\,A_i^* = (1/4)\,p^*\tau(d^*)$ can be written as

$$\frac{\mathrm{d}\Pi_i^*}{\mathrm{d}z} = \frac{1}{4}\left[\frac{\mathrm{d}p^*}{\mathrm{d}z}\tau(d^*) + \tau'(d^*)\frac{\mathrm{d}d^*}{\mathrm{d}z}p^*\right]. \tag{B.9}$$

We look at the effects of advertiser competition intensity. For $z = t_a$ both terms on the

[23] Note that targeting is a function, which we treat as fixed here, such that comparative statics are performed as described in footnote 22.

right-hand side are positive and hence $\mathrm{d}\Pi_i^*/\mathrm{d}t_a > 0$. Intuitively, when competition for advertisers becomes more intense (t_a decreases), then prices for ad-placing decrease. In turn, less data is collected from users, such that targeting becomes less effective, and less total revenue is made on the ad market. Both these effects decrease platform profits.

The intensity of user-side competition increases platforms' surplus, i.e. $\mathrm{d}\Pi_i^*/\mathrm{d}t_u < 0$. This effect is discussed in the main text in Section 5.

Increased nuisance (higher $z = \nu(d)$) increases platforms' surplus, i.e. $\mathrm{d}\Pi_i^*/\mathrm{d}\nu(d) > 0$. More data is collected, which increases targeting and hence the (residual) value of a placed ad, thus also higher prices can be sustained. Overall, this unambiguously benefits platforms.

Increased targeting (higher $z = \tau(d)$) increases platforms' surplus, i.e. $\mathrm{d}\Pi_i^*/\mathrm{d}\tau(d) > 0$. Although less data is collected, the absolute externality of users, i.e. targeting, increases the value to be shared between platforms and advertisers. While the effect on prices remains ambiguous, overall, platforms benefit. To see that note that

$$\frac{\mathrm{d}\Pi_i^*}{\mathrm{d}\tau(d)} = \frac{\tau(d^*) - t_a}{(t_u + \nu(d^*))\,\Psi(d^*)} \left[-\left(t_u + \nu(d^*)\right)\nu'(d^*)\tau'(d^*) \right. \\ \left. +\nu(d^*)\left\{(\tau(d^*) - t_a)\left[\kappa''(d) + \nu''(d^*)\right] - t_c\left(t_u + \nu(d^*)\right)\tau''(d^*)\right\}\right] > 0, \tag{B.10}$$

where $\mathrm{d}d^*/\mathrm{d}t_u$ is from equation (15), while $\Psi(d^*)$ is defined in equation (A.6).

Effects on advertiser profits

The effects on advertiser profits $\pi_i^*(a) = (1-p^*)\tau(d^*)X_i^* - t_a|l_i - a| = (1/2)\,(1-p^*)\,\tau(d^*) - t_a|l_i - a|$ are given by

$$\frac{\mathrm{d}\pi_i^*(a)}{\mathrm{d}z} = \frac{1}{2}\left[-\frac{\mathrm{d}p^*}{\mathrm{d}z}\tau(d^*) + \tau'(d^*)\frac{\mathrm{d}d^*}{\mathrm{d}z}\,(1-p^*)\right] - |l_i - a|\frac{\mathrm{d}t_a}{\mathrm{d}z}. \tag{B.11}$$

Stronger competition for advertisers (lower $z = t_a$) makes advertisers overall better off, i.e. $\mathrm{d}\pi_i^*/\mathrm{d}t_a < 0$. However, there are multiple effects at work. Firstly, prices fall, such that the first term on the right hand side increases. Secondly, less personal data from users can be collected, which makes targeting less effective, therefore the second term is negative. Thirdly, also transportation costs decrease, which increases advertiser profits. Overall, the price and transportation cost reduction effects outweigh decreased targeting effectiveness.

For this note that

$$\begin{aligned}\frac{\mathrm{d}\pi^A}{\mathrm{d}t_a} &= \frac{1}{4\left[t_u+\nu(d^*)\right]^2}\Bigg\{-6t_u\nu(d^*)-\nu(d^*)^2\left[1+2\tau'(d^*)\frac{\mathrm{d}d^*}{\mathrm{d}t_a}\right]\\ &\qquad + t_c\left[-4\nu'(d^*)\frac{\mathrm{d}d^*}{\mathrm{d}t_a}\left(\tau(d^*)-t_a\right)+t_c\left(-5+2\tau'(d^*)\frac{\mathrm{d}d^*}{\mathrm{d}t_a}\right)\right]\Bigg\}\\ &= -\frac{1}{4\left[t_u+\nu(d^*)\right]\Psi(d^*)}\Bigg\{-\nu'(d^*)\left(t_u+\nu(d^*)\right)\left(\tau(d^*)-t_a\right)\tau'(d^*)+3\left(t_u+\nu(d^*)\right)^2\tau'(d^*)^2\\ &\qquad -\left(5t_u+\nu(d^*)\right)\left(\tau(d^*)-t_a\right)\left[-\nu''(d^*)\left(\tau(d^*)-t_a\right)+\left(t_u+\nu(d^*)\right)\tau''(d^*)\right]\Bigg\}\\ &< 0, \end{aligned} \tag{B.12}$$

where $\mathrm{d}d^*/\mathrm{d}t_a$ is from equation (14), while $\Psi(d^*)$ is defined in equation (A.6).

Stronger competition for users (increase $z = t_u$) hurts advertisers, hence $\mathrm{d}\pi_i^A/\mathrm{d}t_u > 0$. The platforms' bottleneck position allows them to increase prices (negative first term) and, further, less user data can be collected, such that targeting becomes less effective (negative second term).

Increased nuisance (higher $z = \nu(d)$) decreases advertisers' surplus, i.e. $\mathrm{d}\pi_i^A/\mathrm{d}\nu(d) < 0$. Although more data is collected, which increases targeting and hence the value of a placed ad, also prices increase. Overall, this hurts advertisers. To see that, note

$$\begin{aligned}\frac{\mathrm{d}\pi^A}{\mathrm{d}\nu(d)} &= -\frac{\left[\tau(d^*)-t_a\right]}{2\left[t_u+\nu(d^*)\right]^2\Psi(d^*)}\Bigg\{\left(t_u+\nu(d^*)\right)^2\tau'(d^*)^2\\ &\qquad +2\left[\tau(d^*)-t_a\right]t_c\left\{\left(\tau(d^*)-t_a\right)\left[\kappa''(d)+\nu''(d^*)\right]-t_c\left(t_u+\nu(d^*)\right)\tau''(d^*)\right\}\Bigg\}\\ &< 0, \end{aligned} \tag{B.13}$$

Increased targeting (higher $z = \tau(d)$) has an ambiguous effect on advertisers' surplus. While the targeting function becomes better, less data needs be collected which again reduces targeting effectiveness. Further, the effect on prices is ambiguous. Hence, overall effects on advertiser surplus remain unclear.

Effects on user utility

The effects on a user's utility $u_i^*(x) = \underline{u} - \kappa(d^*) - \nu(d^*)A_i^* - t_u|l_i - x| = \underline{u} - \kappa(d^*) - (1/2)\,\nu(d^*) - t_u|l_i - x|$ are given by

$$\frac{u_i^*(x)}{\mathrm{d}z} = -\frac{\mathrm{d}d^*}{\mathrm{d}z}\left[\kappa'(d^*) + \frac{\nu'(d^*)}{2}\right] - \frac{\mathrm{d}t_u}{\mathrm{d}z}|l_i - x|. \tag{B.14}$$

Note that by Corollary 1 the term in brackets on the right-hand side is positive and that for $z \in \{t_a, t_u\}$ we have $\mathrm{d}d^*/\mathrm{d}z > 0$ such that $\mathrm{d}u_i/\mathrm{d}z < 0$.

Intuitively, less competition for advertisers (higher $z = t_a$) increases the amount of data collected in equilibrium, which overall leaves users worse off, as privacy concerns are increased, although ads are more targeted and hence nuisance smaller.

Less competition for users (higher $z = t_u$) increases the amount of data collected, such that privacy concerns are increased, although it reduces nuisance costs. Further strengthened by increased transportation costs for users, quite naturally users' utility overall decreases.

Increased nuisance (higher $z = \nu(d)$) decreases users' utility, i.e. $\mathrm{d}u_i/\mathrm{d}\nu(d) < 0$ because again more data is collected.

Increased targeting (higher $z = \tau(d)$) increases users' utility, i.e. $\mathrm{d}u_i/\mathrm{d}\tau(d) < 0$. Although targeting does not directly affect users, less data is collected, which is beneficial for users.

B.3 Market coverage and multi-homing

Advertiser side

We start this section by lifting Assumption 1 for full market coverage and the single-homing assumption for advertisers. Analytically, this is achieved by pinning down advertisers which are indifferent between joining a platform and abstaining such that the total mass of advertisers joining platform i is determined by $\pi_i(a) = 0$.

Figure 2 shows two potential outcomes of this alternative setup. In the first case the total mass of participating advertisers in the market is smaller than 1 while advertisers 'in the middle' choose not to participate as their transportation costs are too high. In the second case the sets of advertisers joining platform i and j are overlapping such that advertisers 'in the middle' join both platforms, i.e. they multi-home. The remaining analysis follows the steps from the baseline model and is omitted at this point.

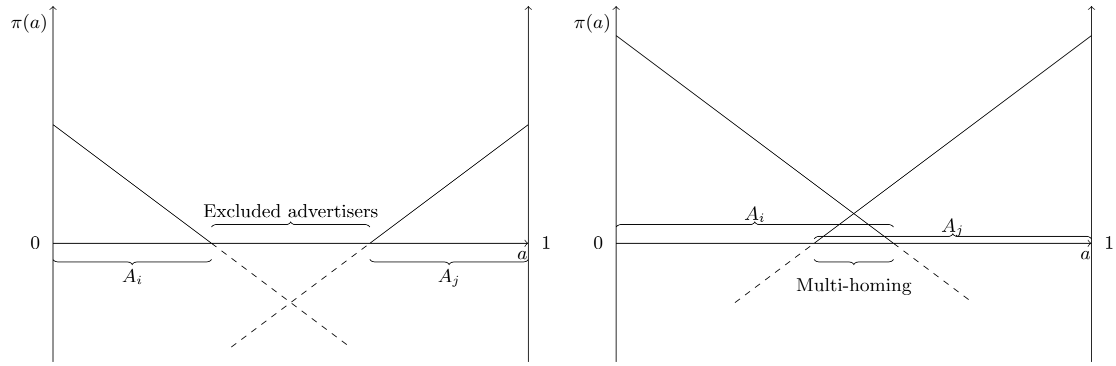

Figure 2: Relaxed advertiser market assumption

The welfare maximizing level of data d_a^o is then given by

$$\kappa'(d_a^o) = A_i^o(d_a^o)\tau'(d_a^o) - A_i^o(d_a^o)\nu'(d_a^o) \tag{B.15}$$

where $A_i^o(d)$ denotes the symmetric mass of advertisers on each platform and is given by $A_i^o(d) = [\tau(d) - \nu(d)]/(2t_a)$. The equilibrium level of data under platform competition d_a^* is then given by

$$\kappa'(d_a^*) = \left(A_i^*(d_a^*)\frac{\nu(d_a^*)}{\tau(d_a^*)} + \frac{t_u}{\tau(d_a^*)} \right) \tau'(d_a^*) - A_i^*(d_a^*)\nu'(d_a^*) \tag{B.16}$$

while $A_i^*(d_a^*) = [(1-p_a^*(d_a^*))\tau(d_a^*)]/(2t_a)$. We can see immediately that whether the resulting allocation is an equilibrium with multi-homing or with excluded advertisers depends on functional forms and parameters. We will therefore discuss the two cases separately in the following.

Assume transport costs t_a are sufficiently low to allow a multi-homing allocation of advertisers under the efficient benchmark, i.e. $A_i^o(d_a^o) > 1/2$. Comparing the condition for the resulting efficient level of data provision to our baseline condition in (12) we see that $d_a^o > d^o$, under multihoming the efficient level of data provision is higher than under single-homing. The idea is that additional advertisers are attracted in order to maximize total value creation in the economy. The comparison of the new competitive level of data provision d_a^* to the new efficiency benchmark as well as to our baseline model is, however, ambiguous. As competition for advertisers is now relaxed, platforms might not be forced to offer high levels of d to attract additional advertisers. At the same the value creation aspect from a larger total number of advertisers is still valid, such that the net effect on the level of data provision remains ambiguous.

When transportation costs t_a are sufficiently large, some advertisers 'in the middle' would not join any platform, such that $A_i^o(d_a^o) < 1/2$ and also $A_i^*(d_a^*) < 1/2$. Note that the efficient level is then also lower than in our benchmark $d_a^o < d^o$ as attracting advertisers becomes relatively expensive and it becomes more efficient to exclude some advertisers than to offer very high levels of d. The comparison to the market outcome, however, remains ambiguous. While the same efficiency argument applies, platforms also have an additional incentive to increase their intensive margin by increasing d to offset the reduction in advertising demand. Again, depending on functional forms either effect may dominate.

User side

Similarly on the user side, by relaxing Assumption 2 it is possible that $\underline{u}$ becomes sufficiently small relative to transportation costs, such that users 'in the middle' prefer to abstain from both platforms. If $\underline{u}$ is sufficiently large relative to transportation costs, users 'in the middle' might choose to join both platforms. In both cases user market shares are determined through the utility of the indifferent user relative to the outside option.

The symmetric welfare-maximizing level of data d_u^o is then given by

$$\kappa'(d_u^o) = X_i^o(d_u^o)\frac{t_u}{\tau(d_u^o) + 2\underline{u} - 2\kappa(d_u^o) - \nu(d_u^o)}\tau'(d_u^o) - \frac{1}{2}\nu'(d_u^o), \tag{B.17}$$

where $X_i^o(d_u^o)$ denotes the symmetric mass of users on each platform and is given by $X_i^o(d_u^o) = [2\underline{u} - 2\kappa(d_u^o) - \nu(d_u^o)]/(2t_u)$. The equilibrium level of data under platform competition d_u^* is then given by

$$\kappa'(d_u^*) = X_i^*(d_u^*)\frac{t_u}{\tau(d_u^*)}\tau'(d_u^*) - \frac{1}{2}\nu'(d_u^*), \tag{B.18}$$

while $X_i^*(d_u^*) = [2\underline{u} - 2\kappa(d_u^*) - \nu(d_u^*)]/(2t_u)$. From this we can immediately see that $d_u^* > d_u^o$, i.e. there is always over-provision of user data. While the efficient benchmark takes into account the tradeoff between excluding users and total value creation, the market outcome only compares the targeting benefit to the exclusion of users. Further note that if the market is not covered such that $X_i(d_u) < 1/2$, the efficient as well as the equilibrium level of data provision is lower than in the baseline model, i.e. $d_u^o < d^o$ and $d_u^* < d^*$ because $t_u/\tau(d) < \delta(d)\ \forall d$.

It is worthwhile to note that under user multi-homing as well as under relaxed user market coverage we get that $\mathrm{d}d_u^*/\mathrm{d}t_u = \mathrm{d}d_u^*/\mathrm{d}t_a = 0$, i.e. the transportation cost parameters on either market side are irrelevant for the equilibrium (and also the efficient) level of data collection. This is because t_u now merely scales demand while the relevant trade-off for the choice of d involves the actual utility from joining the platform, which is unaffected by the demand scale.

Under this setup user demand becomes more elastic than in the baseline model which undermines platforms' incentive to increase d. At the same time platforms would also increase prices $\mathrm{d}p_u^*/\mathrm{d}t_u > 0$ if it becomes increasingly difficult to attract users. Note that we seemingly found the opposite effect in our baseline model $\mathrm{d}p^*/\mathrm{d}t_u < 0$, however, the interpretation of t_u changes substantially such that the two results do not contradict each other: the harder it is to keep users, the higher the prices for advertisers.

In fact platforms are able to overcompensate the reduction in user demand such that $\mathrm{d}\Pi_u^*/\mathrm{d}t_u > 0$ (and for advertisers $\mathrm{d}\pi_u^*/\mathrm{d}t_u < 0$). Again, as the interpretation of t_u essentially reverses, we had the opposite results in our baseline model where platform profits decreased in t_u (while advertiser profits increased). This is also reflected in the effect on the advertiser side where equilibrium prices rise in t_a under both model specifications, i.e. $\mathrm{d}p_u^*/\mathrm{d}t_a > 0$ as the interpretation remains identical.

B.4 Positive cross-group externalities

Consider the following modification of the users' utility function:

$$u_i(x) = \underline{u} - \kappa(d_i) + \rho(d_i)A_i - t_u|l_i - x|. \tag{B.19}$$

The concave and twice-differentiable function $\rho(d)$ represents the relevance from a user's point of view of seeing A_i offers, where $\rho'(d) \geq 0$ and $\rho''(d) \leq 0$. However, $\rho(d)$ can now be entirely negative, positive or might even switch signs. The first case is discussed in depth in the main paper, where we consider the case $\rho(d) = -\nu(d)$. The second case, a strictly positive effect, can be thought of as a traditional 'dating' model, where one group strictly enjoys the presence of the other group. The last case can be thought of as a more nuanced version of our nuisance cost in the baseline model. While for low values of d, i.e. the platform has very little information about the consumer, a user dislikes the interaction with the other market side, the interaction might turn out to be valuable once the platform has sufficient information, i.e. d is sufficiently large. A typical example would be the recommendation system on Amazon. While it is debatable, whether Amazon is a two-sided market in the traditional sense, the product recommendation system might serve as a useful example. A new customer might see all kind of product recommendations, some of which are completely useless to the user and are just a waste of attention. However, once Amazon has acquired sufficient information about the user's preferences through analyzing the purchasing and browsing history, the recommendations become more personalized, and the user finds actual value in looking through them.

From a modelling perspective we only require that the relevance is monotonically increasing in the amount of data, but with decreasing returns. Since the curvature of the maximization problem therefore remains unchanged, the characterization of the second order conditions given in the Appendix A also remain qualitatively unchanged. The absolute value of the function $\rho(d)$ is in the end of minor importance regarding the key mechanics of the model, however, it has to be taken care of through appropriately adjusting the modelling assumptions. In order to assure full market coverage on the offer side, we now have the following set of assumptions.

Assumption 3 *Competition for advertisers is sufficiently strong, i.e.* $t_a \leq \bar{t}_a$.
For this, it is necessary that competition for users is sufficiently weak and that there are gains of trade for all advertisers, even without data collection, i.e.

$$(a) \quad t_u > |\rho(0)|, \rho(d) < t_u$$

$$(b) \quad t_a < \tau(0)$$

The upper bound on t_a is then given by $\bar{t}_a := \frac{t_u\tau(0)+\rho(0)\tau(0)}{3t_u-\rho(0)}$. Since now net cross-group externalities might be positive, a problem of platform tipping must be taken into account. In particular the following assumption ensures that the competitive symmetric equilibrium leads to positive prices (and therefore positive platform profits), so that a platform would not be indifferent whether to enter the market if just one platform serves the entire market.

Assumption 4 *To ensure market participation of both platforms it is necessary to have*

$$t_a t_u > \rho(\cdot)\tau(\cdot).$$

Note that for negative $\rho(\cdot)$ as in our main model, this assumption is always fulfilled as then the RHS is always negative, while the LHS is always positive. Accordingly, if $\rho(\cdot)$ switches signs, the range in which $\rho(\cdot)$ is negative is unproblematic. Therefore the only potentially problematic case is if $\rho(\cdot)$ is positive or can turn positive since it further restricts the parameter space in addition to the previous assumption.[24] Given that both assumptions are satisfied, the analysis is analogous to our main model and all major results still hold.

[24] In the following we sketch a set of conditions under which both assumptions would be satisfied. Note Assumption 4 specifies a lower bound $t_a > \underline{t_a}$ with $\underline{t_a} \equiv \frac{1}{t_u}\rho(\cdot)\tau(\cdot)$. It is therefore necessary to show that the set of t_a satisfying Assumptions 3 and 4 is non-empty. In particular, if it holds that $lim_{d \to \infty} \underline{t_a} < \bar{t_a}$ we can always find intermediate values of t_a satisfying both conditions. For this to be the case it is necessary that $lim_{d \to \infty} \underline{t_a} < \tau(0)$ and that $\rho(\cdot)$ is small if positive.