Privacy and Platform Competition

Philipp Dimakopoulos, Humboldt-Universität zu Berlin
Slobodan Sudaric, Humboldt-Universität zu Berlin
Privacy and Platform Competition

Philipp Dimakopoulos
Humboldt-Universität zu Berlin

Slobodan Sudaric
Humboldt-Universität zu Berlin

August 7, 2017

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 effects on both market sides. Moreover, if platforms engage in two-sided pricing, data provision would be efficient.

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

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

The authors would like to thank Pio Baake, Helmut Bester, Colin von Negenborn, Markus Reisinger, Roland Strausz, Jonas von Wangenheim as well as participants of the CRESSE on Crete, IODE in Liège, MaCCI in Mannheim, BCCP in Berlin and seminar participants in Berlin for helpful comments and suggestions. 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 consumers 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 consumers 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 consumer side. Consumers 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 consumer perspective can be interpreted as a price the consumer is willing to accept in exchange for the use of the platform including the display of ads.\footnote{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 consumers are aware and willing to exchange personal data for services, however, the actual extent worries them.} 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).\footnote{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).} And indeed, recent cases demonstrate that competition authorities acknowledge the peculiarities of data-driven industries. Germany’s Federal Cartel Office (Bundeskartellamt, BKA) initiated investigations against Facebook in 2016 based on alleged abuse of market power. In particular, the BKA investigates whether Facebook

\begin{itemize}
  \item[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 consumers are aware and willing to exchange personal data for services, however, the actual extent worries them.
  \item[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).
\end{itemize}
uses its dominant position in the market for social networks to expand the terms of service outlining how much data is collected and processed by the platform.\textsuperscript{3} We therefore 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.\textsuperscript{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 platforms have significant market power, 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. Our results indicate that the inefficiency of data provision can be reduced


\textsuperscript{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.
by careful privacy regulation or competition policies on either market side. Although competition policy measures on both market sides are substitutable, there is a difference in effectiveness: in a situation of overprovision of user data, it is more effective to strengthen platform competition on the advertiser side than on the user side, and vice-versa. One interpretation of this result is that 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 consumer side, and allow for non-zero consumer prices. In fact, lifting the restriction leads to an efficient level of collected data, while consumer 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 consumer side leads to an efficient outcome. This reflects the idea of Laudon (1996) that consumers 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. Secondly, this result gives rise to a counterfactual interpretation. In particular we argue that whenever the unrestricted model would yield positive (negative) consumer prices, the restricted model exhibits overprovision (underprovision) of user data as platforms can no longer adequately charge or compensate consumers for collecting data. Secondly, we also demonstrate how our model can be used to analyze cases with positive cross-group externalities (matching platforms) and in particular allow for the case where consumers obtain disutility from interacting with the other market side (pure nuisance), but the interaction turns positive once the platform collected a sufficient amount of data. Lastly, we sketch the case of platform collusion and show that the amount of collected data in this case is excessively high which stresses the importance of effective competition including non-price dimensions.

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 the baseline model with respect to two-sided pricing, positive cross-group externalities and collusion. Section 8 concludes.
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 consumers but consumers 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 consumer as well as perceived nuisance are constant. Our research differs as 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 consumers 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 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 the economic aspects of privacy. While a big part of the literature is focused on how consumer information can be used for price discrimination (see e.g. Villas-Boas, 2004; Taylor, 2004; Acquisti and Varian, 2005; Conitzer et al., 2012) and ad-avoidance (e.g. Anderson and Gans, 2011; Hann et al., 2008, Johnson, 2013), we want to shed light on the efficient provision of personal data, and therefore relate our paper to the literature on efficient privacy provision and the role of privacy as a competition instrument. The aspect of data provision being a strategic decision
made by platforms is captured to some extent by Spiegel (2013). The question here is how commercial software (full privacy) compares to adware (positive privacy costs), i.e. software which is financed through targeted advertising. He shows that adware is welfare superior to the provision of full-privacy commercial software. De Corniere and De Nijs (2016) consider a setting where a monopolistic platform auctions off advertising slots and decides whether to disclose information regarding potential consumers (no privacy) or not (privacy). They show that in particular if the number of bidders is large, the platform prefers information disclosure. However, this comes at the cost of some consumers leaving the market as equilibrium consumer prices rise such that from a welfare point of view it is not clear which regime is preferable. Also, 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. As the revenue side of the market is not modelled explicitly, the trade-off faced by the platform boils down to an intensive vs. extensive margin problem and whether the platform’s incentives are aligned with consumer interests depends crucially on the elasticity of consumer demand. As the mentioned papers consider the case of monopolistic platforms, we see our research as a natural extension since we consider the case of competing platforms and allow 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. In fact, if consumers have a low willingness to pay for the product increased competition intensity can lead to less privacy. 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 consumer 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 consumers 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 by the firm’s customers 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 consumers. Advertisers and consumers are distributed uniformly on different Hotelling lines of unit length and are assumed to both single-home. While we acknowledge that this is a strong assumption it allows us to focus on the role of competition intensity more clearly. 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 competitive environment and hence the relevant market of platforms when competing for users may be different from the market when competing for advertisers. For example, different online platforms, like 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 $i$, where

$$u_i(x) = u - \kappa(d_i) - \nu(d_i)a_i - t_u|l_i - x|.$$  \hspace{1cm} (1)

The first term of the utility function is a fixed utility component $u$ 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 non-concavity on these costs, i.e. $\kappa'(d_i) \geq 0$ and $\kappa''(d_i) \geq 0$. Third, users incur nuisance cost $\nu(d_i) \leq 0$ per advertisements $a_i$ on the platform. These nuisance costs fall in the amount of provided data $d_i$ as we assume that users prefer personalized to non-personalized ads, i.e. $\nu'(d_i) < 0$ and $\nu''(d_i) > 0$. This setup reflects the idea, that the more relevant an ad, the higher the chance of value creation through a possible follow-up purchase.\(^5\) Finally, users face transportation costs due to horizontal platform differentiation, whereby we assume uniform user distribution on the Hotelling line, i.e. $x \sim [0, 1]$, while $t_u > 0$ is the associated transportation cost parameter.

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 consumers 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 consumers, 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

\(^5\)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.
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 $i$, where

$$\pi_i(a) = \tau(d_i)(1 - p_i)x_i - t_a|l_i - a|. \quad (2)$$

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 function $\tau(d_i) \geq 0$ can be interpreted as the targeting ability of platforms: the more data $d_i$ 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 $\tau'(d_i) > 0$ and $\tau''(d_i) < 0$. 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. $a \sim [0, 1]$, 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 ($u$) to consumers, revenue is only generated through presenting ads to consumers. Platform profits are then given by

$$\Pi_i(d_i, p_i) = a_i\tau(d_i)p_ix_i \quad (3)$$
where \( a_i \) denotes the number of advertisers at platform \( i \) and \( p_i \) is the per-click price advertisers have to pay if \( x_i \) users click with probability \( \tau(d_i) \).\(^6\) 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 a crucial input factor for the click-probability the advertisers are facing. At the same 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 assumption to assure full advertiser market coverage.

**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.

\[
\begin{align*}
(a) \quad & t_u > \nu(0) \\
(b) \quad & t_a < \tau(0)
\end{align*}
\]

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 \( t_a \) is merely technical, allowing us to isolate effects in a competitive environment. Intuitively, this constitutes a sufficient condition, such that for any level of (symmetric) data provision \( d \geq 0 \), it is assured that advertisers have enough market power to obtain positive profits such that their market is fully covered. Consequently, competition for advertisers in 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 \( t_u > \nu(0) \).\(^7\) 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 \( a_j = 1 \), at least the user most

---

\(^6\)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.

\(^7\)Note that \( t_u > \nu(0) \Rightarrow t_u > \nu(d) \forall d \) because \( \nu'(d) < 0 \).
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.

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 consumers 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 \( u \) is large enough to ensure full participation on the user side.

Intuitively, the platform service provides a utility high enough such that consumers are not deterred through the provision of personal data and the incurred nuisance from advertising.

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 consumers observe the platforms’ choices and simultaneously decide which platform to join, hence determining \( a_i \) and \( x_i \). 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 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

---

\(^8\) We could also consider an alternative timing where advertisers choose first and consumers last. The outcome is equivalent in our model.
user market coverage due to Assumption 2, the number of consumers $x_i$ joining platform $i$ is then determined by the indifferent consumer $\hat{x}: u_i(\hat{x}) = u_j(\hat{x})$ such that

\[ x_i = \hat{x} = \frac{1}{2} + \frac{1}{2t_u} [\kappa(d_j) - \kappa(d_i) + \nu(d_i)a_i - \nu(d_j)a_j] \quad (4) \]

\[ x_j = 1 - \hat{x} = \frac{1}{2} + \frac{1}{2t_u} [\kappa(d_i) - \kappa(d_j) + \nu(d_i)a_i - \nu(d_j)a_j] \quad (5) \]

Similarly, market shares on the advertiser side are given by the indifferent advertiser $\hat{\alpha}: \pi_i(\hat{\alpha}) = \pi_j(\hat{\alpha})$. 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

\[ a_i = \hat{\alpha} = \frac{1}{2} + \frac{1}{2t_a} [\tau(d_i)(1-p_i)x_i - \tau(d_j)(1-p_j)x_j] \quad (6) \]

\[ a_j = 1 - \hat{\alpha} = \frac{1}{2} + \frac{1}{2t_a} [\tau(d_j)(1-p_j)x_j - \tau(d_i)(1-p_i)x_i] \quad (7) \]

Solving the system of equations given in (4) - (7) gives us the following result.

**Lemma 1** Second stage equilibrium market shares are given by

\[ x_i(d_i,d_j,p_i,p_j), x_j(d_i,d_j,p_i,p_j) \text{ and } a_i(d_i,d_j,p_i,p_j), a_j(d_i,d_j,p_i,p_j). \quad (8) \]

### 4.2 Efficiency Benchmark

Let us start with deriving the welfare efficient benchmark. For this we define welfare as the sum of all indirect utilities and profits, anticipating second stage market shares as in Lemma 1, i.e.

\[ W(d_i,d_j,p_i,p_j) = \int_0^{x_i} u_i(x)dx + \int_{x_i}^1 u_j(x)dx + \int_{a_i}^{a_i} \pi_i(a)da + \int_{a_j}^{a_j} \pi_j(a)da + \Pi_i + \Pi_j. \quad (9) \]

**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

\[
\kappa'(d^o) = \frac{\tau'(d^o)}{2} - \frac{\nu'(d^o)}{2} \tag{10}
\]

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.

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.\(^9\) 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.\(^10\)

**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 (9), anticipating second stage market shares as in Lemma 1.\(^11\)

**Proposition 2**  User utility is maximized by the unique symmetric solution \((d^u, p^u) = (d^u_i, p^u_i)\) for \( i \in \{1, 2\} \), where \( d^u \) is characterized by

\[
\kappa'(d^u) = -\frac{1}{2} \nu'(d^u_i), \tag{11}
\]

\(^9\)Note 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.

\(^10\)Note that 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 (14).

\(^11\)Note that the same data level \( d^u \) would result if we only choose \( d_i \) to maximize consumer surplus, while anticipating firms setting ad prices \( p_i \) subsequently. These prices would be identical to the prices in the market outcome, given by equation (14).
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. $a^u_i = 1/2$ and $x^u_i = 1/2$.

4.4 Market Outcome

For the market outcome, in the first stage platforms maximize their profits, anticipating second stage market shares as in Lemma 1.

$$\max_{p_i, d_i} \Pi_i (d_i, p_i) = a_i (d_i, d_j, p_i, p_j) \tau(d_i) p_i x_i (d_i, d_j, p_i, p_j) \quad \forall i \in \{1, 2\}$$

(12)

and we obtain a symmetric solution for prices and data levels from the first order conditions. Regarding the curvature of the maximization problem we can say that the solution to the first-order condition 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 the appendix.

**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 user is implicitly given by

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

(13)

and prices per advertisement are

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

(14)

resulting in equal advertiser as well as user distribution to both platforms, i.e.

$$a^*_i = \frac{1}{2} \text{ and } x^*_i = \frac{1}{2}.$$  

(15)

Comparing the market level of data provision $d^*$ in (13) to the efficient level $d^j$ in (10) 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.

Then, in this equilibrium, we get the following advertiser profit $\pi^*_i(a)$, user utility $u_i^*(x)$
and platform profits $\Pi^*_i$.  
\[
\pi^*_i(a) = \frac{\tau(d^*)}{2} - \frac{t_at_u + \nu(d^*)\tau(d^*)}{t_u + \nu(d^*)} - t_a \min \{a, 1 - a\} \quad (16)
\]
\[
u^*_i(x) = u - \kappa(d^*) - \frac{\nu(d^*)}{2} - t_u \min \{x, 1 - x\} \quad (17)
\]
\[
\Pi^*_i = \frac{t_at_u + \nu(d^*)\tau(d^*)}{2[t_u + \nu(d^*)]} \quad (18)
\]

Note that with a sufficiently high baseline utility $u$, user utility will always be non-negative.

Further note that the equilibrium price $p^*$ does not exceed one and that advertiser profits as given by equation (16) will be positive for all advertisers due to assumption 1. The following corollary summarizes this finding.

**Corollary 1** In equilibrium, $p^* < 1$ and $\pi^*_i(a) \geq 0$.

**Proof.** Given equation (14), $p^* < 1$ if
\[
2 \cdot \frac{t_at_u + \nu(d^*)\tau(d^*)}{\tau(d^*)t_u + \nu(d^*)\tau(d^*)} < 1 \iff t_a < \tau(d^*) \frac{(t_u - \nu(d^*))}{2t_u} < \tau(d^*) \quad (19)
\]

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
\[
\pi^*_i\left(\frac{1}{2}\right) = \frac{\tau(d^*)}{2} - \frac{t_at_u + \nu(d^*)\tau(d^*)}{t_u + \nu(d^*)} - \frac{t_a}{2} \geq 0 \iff \tau(d^*) \frac{t_u - \nu(d^*)}{3t_u + \nu(d^*)} \geq t_a, \quad (20)
\]

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 $d$. ■

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

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

**Proof.** Rearranging terms in the first-order condition of platform profit maximization,
Figure 1: Overview of comparative statics

<table>
<thead>
<tr>
<th></th>
<th>( \frac{d\kappa}{dz} )</th>
<th>( \frac{dp}{dz} )</th>
<th>( \frac{d\Pi^*_i}{dz} )</th>
<th>( \frac{d\pi^*_i}{dz} )</th>
<th>( \frac{du^*_i}{dz} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_a )</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( t_u )</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>( \nu(d) )</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \tau(d) )</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>+</td>
<td>?</td>
</tr>
</tbody>
</table>

given by equation (13), yields

\[
2\kappa'(d^*) + \nu'(d^*) = \tau'(d^*) \frac{\nu(d^*) + t_u}{\tau(d^*) - t_a} > 0. \tag{21}
\]

By assumption 1 we have \( \tau(d^*) > t_a \). Hence the second term on the right hand side of equation (21) must be positive, such that \( 2\kappa'(d^*) + \nu'(d^*) > 0 \).

Intuitively, Corollary 2 says 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 compared to the user-optimal level is provided.\(^{12}\)

5 Comparative statics

In this section we want to provide economic intuition for the equilibrium results of our model. For this we will assess 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 targeting \( \tau(d) \) on equilibrium values of personal data provision \( d^* \), ad-per-click price \( p^* \), as well as platform profits \( \Pi^*_i \), advertiser profits \( \pi^*_i \) and user utility \( u^*_i \).

The table in Figure 1 provides an overview of the derived comparative statics results while the remaining section provides detailed derivations (with some proofs in Appendix A.2, whenever necessary) and intuition, before we turn to policy implications in section 6.

\(^{12}\) In section 6 we provide more details on the comparison of the market outcome with the user-optimal outcome.
5.1 Effects on prices and collected data in equilibrium

Initially, from differentiating equation (14) note that the more data is collected in equilibrium, the lower the ad price:

\[
\frac{dp^*}{dd^*} = 2\frac{\nu'(d^*) \tau(d^*) (\tau(d^*) - t_a) t_u - \tau'(d^*) (t_u + \nu(d^*)) t_a t_u}{[t_u + \nu(d^*)]^2 (\tau(d^*))^2} < 0,
\]

This gives rise to the following proposition.

**Proposition 4** Collecting data from users or money from advertisers are substitutes for the platforms.

Platforms charge money from advertisers and data from users. Both inputs increase platforms’ profits, ceteris paribus. In equilibrium, these inputs are substitutes for the platforms, in the sense that if they collected marginally more data, they would in turn marginally reduce the ad price. This is because collecting more data would repel users, thereby decreasing advertiser demand, which could only be held up by reducing ad prices. Intuitively, in the competitive equilibrium platforms cannot further increase the data ‘payment’ without having to adjust the ad price payment respectively.

Subsequently, in this subsection we evaluate how the intensity of competition as well as the intensity of nuisance and targeting effectiveness affect market outcomes in equilibrium. For this we have to distinguish between the platform competition intensity on the user side and on the advertiser side. Since we have horizontally differentiated platforms vis-a-vis users as well as advertisers, 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.

5.1.1 Competition for users

First, we evaluate the effects of user-side competition intensity on data collection. Since we have only implicit solutions for \(d^*\), we make use of the implicit function theorem by totally differentiating the first-order conditions from equations (13) and (14) w.r.t. \(t_u\).
Solving for \( \frac{dd^*}{dt_u} \) yields

\[
\frac{dd^*}{dt_u} = \frac{(\tau(d^*) - t_a) \tau'(d^*)}{\Psi(d^*)} > 0, \tag{23}
\]

where the term in the denominator is given by

\[
\Psi(d^*) = [2\kappa''(d^*) + \nu''(d^*)] (\tau(d^*) - t_a)^2 - \nu'(d^*) \tau'(d^*) (\tau(d^*) - t_a)
\]
\[
+ (\nu(d^*) + t_u) [\tau'(d^*)^2 - (\tau(d^*) - t_a) \tau''(d^*)] > 0,
\tag{24}
\]

as \( \tau'(d^*) > 0 \) and \( \tau''(d^*) < 0 \) while \( \nu'(d^*) < 0 \) and \( \nu''(d^*) > 0 \) by construction, and \( \tau(d^*) > t_a \) by Assumption 1.

Second, we analyze the effects of competition intensity for users on \( p^* \). While we have an explicit solution for \( p^* \), we still need to take into account the second-order effect of \( t_u \) on \( p^* \) through \( d^* \). From equation (14), we get for the derivative of \( p^* \) w.r.t. \( t_u \)

\[
\frac{dp^*}{dt_u} = 2 - \nu'(d^*) \left[ \tau(d^*) - t_a \right] \tau(d^*) - \frac{dd^*}{dt_u} \left[ -\nu'(d^*) \tau(d^*) (\tau(d^*) - t_a) + \tau'(d^*) (t_u + \nu(d^*)) t_u \right] < 0,
\tag{25}
\]

since \( t_a < \tau(d^*) \) by assumption 1 and \( dd^*/dt_u > 0 \) as established above.

The following propositions sums up comparative statics of user-side competition intensity.

**Proposition 5** When platform competition for users intensifies,

- less user data is collected and
- the ad price-per-click for advertisers increases.

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 amount 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 repelling valuable users. Hence, advertiser prices can increase in
equilibrium. These two results somewhat reflect the “standard” two-sided platform logic: stronger competition on one side of the market reduces this side’s “price”, while it increases the other side’s.

5.1.2 Competition for advertisers

First, we consider the effects of advertiser-side competition on data collection. Again we make use of the implicit function theorem by totally differentiating the first-order conditions from equations (13) and (14) w.r.t. $t_a$. Solving for $d^*/dt_a$ yields

$$\frac{dd^*}{dt_a} = \frac{(\nu(d^*) + t_u) \tau'(d^*)}{\Psi(d^*)} > 0.$$  \hfill (26)

Second, we evaluate the effects of competition intensity for advertisers on $p^*$. From equation (14), we get for the derivative of $p^*$ w.r.t. $t_a$

$$\frac{dp^*}{dt_a} = 2t_u \frac{[\nu(d^*) + t_u] \tau(d^*) - \frac{dd^*}{dt_a} [\nu'(d^*) (\nu(d^*) - t_a) \tau(d^*) + \tau'(d^*) (t_u + \nu(d^*)) t_a]}{[t_u + \nu(d^*)]^2 (\tau(d^*))^2}. \hfill (27)$$

Because $\nu'(d^*) < 0$ by construction, $t_a < \tau(d^*)$ by assumption 1 and $dd^*/dt_a > 0$ as established above, both terms in the numerator have opposing signs and we need further analysis. For this, insert $dd^*/dt_a$ from equation (26) into $dp^*/dt_a$ from equation (27), which simplifies to

$$\frac{dp^*}{dt_a} = \frac{-2t_u (\tau(d^*) - t_a) \left[ -\nu''(d^*) \tau(d^*) (\nu(d^*) - t_a) - (\nu(d^*) + t_u) (\tau'(d^*)^2 - \tau(d^*) \tau''(d^*)) \right]}{(\nu(d^*) + t_u) \tau(d^*)^2 \Psi(d^*)} > 0.$$  \hfill (28)

**Proposition 6** When platform competition for advertisers intensifies

- the ad price-per-click for advertisers falls and
- less user data is collected.

Intuitively, lower advertiser transportation costs mean less sticky advertisers and hence increased platform competition for advertisers. Therefore, it is straightforward that advertiser prices fall. At the same time this would increase the share of advertisers on a platform, thereby repelling users. As a consequence platforms will reduce the level of collected user
data, such as not to shy away users. Contrary to the mechanics of comparative statics of user transportation costs, this effect does not follow “standard” two-sided platform logic as here more competition for advertisers reduces users’ data “payment”.

5.1.3 Nuisance

First, we consider the effects of nuisance on data collection.\textsuperscript{13} Totally differentiating the first-order conditions from equations (13) and (14) w.r.t. $\nu(d)$ and solving for $dd^*/d\nu(d)$ yields

$$\frac{dd^*}{d\nu(d)}|_{d=d^*} = \frac{(\tau(d^*) - t_a) \tau'(d^*)}{\Psi(d^*)} > 0. \quad (29)$$

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

$$\frac{dp^*}{d\nu(d)}|_{d=d^*} = \frac{-2tu(t_a - \tau)^2 [\tau(\nu'(t_u) - (\nu - t_u)(2\kappa'' + \nu'')) - (\nu + t_u) \tau'^2]}{(\nu + t_u) \tau^2 \Psi(d^*)} > 0. \quad (30)$$

**Proposition 7** When nuisance intensifies

- the ad price-per-click for advertisers increases and
- more user data is collected.

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.

\textsuperscript{13}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 $dd^*/d\nu(d)|_{d=d^*}$ we intuitively consider the effect of adding a positive constant $c$ to the function, i.e. $\nu(d) + c$ where $c > 0$, on $d^*$.  

20
5.1.4 Targeting

First, we consider the effects of the targeting technology on data collection.\(^{14}\) Solving for \(dd^*/d\tau(d)\) yields

\[
\frac{dd^*}{d\tau(d)} \bigg|_{d=d^*} = -\frac{(\nu(d^*) + t_u) \tau'(d^*)}{\Psi(d^*)} < 0. \tag{31}
\]

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

\[
\frac{dp^*}{d\nu(d)} \bigg|_{d=d^*} = \frac{2t_u (\tau - t_a) \left[ \nu'' (\nu + t_u) t_a - (\tau - t_a) \left[ \nu' \tau' + t_a (2\kappa'' + \nu'') \right] \right]}{(\nu + t_u) \tau^2 \Psi(d^*)} \geq 0. \tag{32}
\]

**Proposition 8** When targeting becomes more effective

- the effect on the ad price-per-click for advertisers is ambiguous and
- less user data is collected.

First, better targeting allows platforms to create the same ad value with less personal data, hence in equilibrium platforms will compete to "relax" the data requirement for users. Second, 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.

5.2 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.

\(^{14}\)Note that targeting is a function, which we treat as fixed here, such that comparative statics are performed as described in footnote 13.
5.2.1 Effects on platform profits

The effects on platform profits $\Pi_i^p = p^* \tau(d^*) a^*_i a^*_i = (1/4) p^* \tau(d^*)$ can be broken down to

$$\frac{d\Pi_i^p}{dz} = \frac{1}{4} \left[ \frac{dp^*}{dz} \tau(d^*) + \tau'(d^*) \frac{dd^*}{dz} p^* \right]. \quad (33)$$

We look at the effects of advertiser competition intensity. For $z = t_a$ both terms on the right-hand side are positive and hence $d(\Pi_i^p)/dt_a > 0$. 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 effect of user-side competition is less straight-forward. For $z = t_u$, the first term on the right-hand side of (33) is negative, while the second term is positive. Intuitively, when competition for users becomes more intense ($t_u$ decreases), less data can be collected from users, which leads to less effective ad targeting, hence the second term is negative. Nevertheless, the bottleneck position of platforms enables them to increase ad prices on the other side. This is the positive first-term effect, which is stronger in equilibrium. Hence, overall, platforms benefit from harsher competition for users, i.e. $d(\Pi_i^p)/dt_u < 0$. This effect might seem counter-intuitive at first sight. However, it is important to note that platform revenues are exclusively made on the advertiser side and also the fact that platforms constitute a bottleneck for user access and therefore for personal data as well.

Increased nuisance (higher $z = \nu(d)$) increases platforms’ surplus, i.e. $d(\Pi_i^p)/d\nu(d) > 0$. 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. $d(\Pi_i^p)/d\tau(d) > 0$. 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.
5.2.2 Effects on advertiser profits

The effects on advertiser profits $\pi^A_i(a) = (1 - p^*) \tau(d^*) x^*_i - t_u |l_i - a| = (1/2) (1 - p^*) \tau(d^*) - t_u |l_i - a|$ are given by

$$\frac{d\pi^A_i(a)}{dz} = \frac{1}{2} \left[ -\frac{dp^*}{dz} \tau(d^*) + \tau'(d^*) \frac{dd^*}{dz} (1 - p^*) \right] - |l_i - a| \frac{dt_u}{dz}. \quad (34)$$

Stronger competition for advertiser (lower $z = t_a$) makes advertisers overall better off, i.e. $d\pi^A_i/a < 0$. This is because, 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.

Stronger competition for user (increase $z = t_u$) hurts advertisers, hence $d\pi^A_i/dt_u > 0$. 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). This effect is in line with the classic platform effect that if one side becomes more price-elastic (here more competitive), then the other side has to pay more.

Increased nuisance (higher $z = \nu(d)$) decreases advertisers’ surplus, i.e. $d\pi^A_i/d\nu(d) < 0$. Although more data is collected, which increases targeting and hence the value of a placed ad, also prices increase. Overall, this hurts advertisers.

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.

5.2.3 Effects on user utility

The effects on a user’s utility $u^*_i(x) = u - \kappa(d^*) - \nu(d^*) a^* - t_u |l_i - x| = u - \kappa(d^*) - (1/2) \nu(d^*) - t_u |l_i - x|$ are given by

$$\frac{u^*_i(x)}{dz} = - \frac{dd^*}{dz} \left[ \kappa'(d^*) + \frac{\nu'(d^*)}{2} \right] - \frac{dt_u}{dz} |l_i - x|. \quad (35)$$
Note that by Lemma 2 the term in brackets on the right-hand side is positive and that for $z \in \{t_a, t_u\}$ we have $dd^*/dz > 0$ such that $du_i/dz < 0$.

More intense competition for advertisers (lower $z = t_a$) increases users’ utility, i.e. $du_i/dt_a < 0$. Intuitively, higher advertiser competition reduces the amount of data collected in equilibrium, which overall leaves users better off, as privacy concerns are reduced, although ads are less targeted and hence nuisance higher.

Stronger user competition (lower $z = t_u$) quite naturally increases users’ utility, i.e. $du_i/dt_u < 0$. Again, less data is collected, which on the one hand reduces privacy concerns and on the other hand increases nuisance costs. Overall, the positive effects prevail and are further strengthened by reduced transportation costs for users.

Increased nuisance (higher $z = \nu(d)$) quite naturally decreases users’ utility, i.e. $du_i/d\nu(d) < 0$. More data is collected, which on the one hand increases privacy concerns and on the other hand (relatively) decreases nuisance costs. Overall, the negative effects prevails in equilibrium as too much data relative to the user optimum is collected in any case.

Increased targeting (higher $z = \tau(d)$) increases users’ utility, i.e. $du_i/d\tau(d) < 0$. Although targeting does not directly affect users, less data needs be collected, which is beneficial for users.

6 Policy Implications

In this section we will draw comparisons between the different outcomes presented in section 4. Further, we will present policy implications from these.

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 (13) and the efficient level $d^0$ in (10) we can see that the difference will crucially depend on the distortion induced by

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

(36)

24
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 \( \delta(d^*) > 0 \) in equilibrium. Graphically, if \( \delta(d^*) \) increases, this is represented in a right shift of the function pinning down the optimal level \( d^* \) as an intersection with \( \kappa'(d) \). Figure 2 shows a graphical representation of the FOCs determining the respective optimal levels of \( d \).

In fact we can see that depending on parameter values there can be underprovision \( (d^*_u < d^0) \) as well as overprovision \( (d^*_o > d^0) \) of personal data in the competitive equilibrium compared to the efficient benchmark. In particular we can infer from equations (13) and (10) 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 (13) simplifies to (10), the efficient level of data provision. Using our definition of \( \delta(d^*) \) we can then see that

\[
\delta(d^*) < 1 \iff \tau(d^*) - \nu(d^*) > t_a + t_u \quad (37)
\]

and \( d^* > d^0 \) if

\[
\delta(d^*) > 1 \iff \tau(d^*) - \nu(d^*) < t_a + t_u \quad (38)
\]

These results are summarized in the two following proposition.
Proposition 9 The competitive outcome leads to overprovision (underprovision) of personal data if competition on both market sides is weak (strong) and/or if net cross-group externalities are small (large).

Proof. The proof relies on the monotonicity of the LHS and RHS in equations (10) and (13). Suppose, \( \delta(d^*) > 1 \) but \( d^* < d^0 \) and hence \( \kappa'(d^*) \leq \kappa'(d^0) \). Using the implicit definition of \( d^0 \) in (10) and \( d^* \) in (13) this implies \( \delta(d^*)\tau'(d^*) - \nu'(d^*) \leq \tau'(d^0) - \nu'(d^0) \).

Rearranging yields \( \delta(d^*) \leq \frac{\tau'(d^0)}{\tau'(d^*)} + \frac{\nu'(d^*)-\nu'(d^0)}{\tau'(d^*)} \). But due to the curvature of \( \tau(\cdot) \), \( \nu(\cdot) \) we have \( \frac{\tau'(d^0)}{\tau'(d^*)} < 1 \) and \( \frac{\nu'(d^*)-\nu'(d^0)}{\tau'(d^*)} < 0 \) for \( d^* < d^0 \), contradicting \( \delta(d^*) > 1 \). Now suppose \( \delta(d^*) > 1 \) and \( d^* > d^0 \), and hence \( \delta(d^*) \leq \frac{\tau'(d^0)}{\tau'(d^*)} + \frac{\nu'(d^*)-\nu'(d^0)}{\tau'(d^*)} \). For \( d^* > d^0 \) we then have \( \frac{\tau'(d^0)}{\tau'(d^*)} > 1 \) and \( \frac{\nu'(d^*)-\nu'(d^0)}{\tau'(d^*)} > 0 \) and hence \( \delta(d^*) > 1 \).

We want to interpret this finding by first holding privacy concerns \( \kappa(d) \) and the functions \( \nu(d), \tau(d) \) fixed and asking the question which competitive environment leads to which scenario. Proposition 9 tells us if competition on both sides is too strong, i.e. \( t_a + t_u \) is small, platforms tend to collect and process an inefficiently small amount of data. Since consumers are likely to switch to more favorable data provision offers, the platforms’ ability to gather data is limited, \( \frac{dd^*}{dt_a} > 0 \). This increase in competition for consumers increases on the one hand the value of the collected data, \( \frac{dp^*}{dt_a} < 0 \). However, if competition for advertisers is strong as well this effect might be offset by competitive pressure, \( \frac{dp^*}{dt_u} > 0 \).

A similar argument can be made if in turn competition on both sides is weak, i.e. \( t_a + t_u \) is high. Consumers are likely to accept higher degrees of data collection due to difficulties switching to a competing platform. The resulting data inflation depresses prices on the advertiser market. However, if market power is sufficiently high the adverse effect can be offset and platforms have a monetary incentive to collect too large amounts of data. To summarize, note that in a scenario of underprovision an increase in competition intensity may have detrimental welfare effects in these markets.

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 \(-\nu(d^*) \). If these effects together are relatively large (small), the LHS of equations (37) and (38) become relatively
large (small) and hence we are in a situation of underprovision (overprovision).
Comparing the user-optimal level $d^c$ to the welfare optimal level $d^o$ we immediately see that users provide an inefficiently low level of data. This result is summarized in the following proposition.

**Proposition 10** The user-optimal level of data provision is inefficiently low.

The reason for this result is straight forward. 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 2. In particular, even in the case of underprovision, the competitive outcome is better from a welfare perspective than the user-optimal choice. This situation is also depicted in graph 2 and summarized in the following corollary.

**Corollary 3** If the market outcome leads to underprovision of personal data, it still outperforms the user-optimal choice in terms of welfare.

Unlike users, platforms act as intermediaries and are able to internalize parts of the value creation on both sides of the market. Depending on the structure of the market this might lead to putting too much or too little weight on the advertiser side of the market, resulting in a situation where the data collection is inefficiently low or high. However, if it turns out that there is an underprovision of personal data, the competitive outcome is closer to the welfare optimal level, since the additional positive effect on the advertiser market side is internalized.\(^{15}\)

### 6.2 Policy Conclusions

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

\(^{15}\)Note even a scenario of overprovision is better in terms of welfare at least up to threshold.
In our model, an omnipotent regulator could obviously achieve the first-best outcome by forcing $d_i = d_j = d^0$ 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 BKKartA or the Facebook/Whatsapp merger case in the US and the EU. In this section we want to present answers our model provides for 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 $d_i = d_j = d^0$. 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 consumer-optimal amount of data is always inefficiently low as consumers 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 9) 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 \((t_u + t_a \text{ low})\), 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.

First of all, 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 \(d^*\), i.e. platforms’ market reaction to competition policies, towards the welfare optimum \(d^o\). This result is summarized in the following corollary.

**Corollary 4** *Advertiser and user side competition policies are substitutes in their effect on platforms’ equilibrium data collection.*

Even though both competition parameters work in the same direction, i.e. more competition yields less data collection, they are not always equally effective. Keeping in mind the implicit definition of \(d^*\) in equation (13) and going back to Figure 2 we can see that shifts in the transportation cost parameters correspond to shifts of the graph in a one-to-one relationship. Since the only source of distortion in our model is an inefficient amount of data, we can ask ourselves which parameter leads to a stronger reaction of \(d^*\). We can therefore look at the reaction of the RHS of equation (13) \(RHS(d^*) := \frac{1}{2} \left[ \frac{\nu'(d^*) + t_u}{\tau(d^*) - t_u} \tau'(d^*) - \nu'(d^*) \right]\) such that

\[
\frac{dRHS(d^*)}{dt_a} = \frac{1}{2} \frac{\nu(d^*) + t_u}{(\tau(d^*) - t_a)^2} \tau'(d^*) \\
\frac{dRHS(d^*)}{dt_u} = \frac{1}{2} \frac{1}{\tau(d^*) - t_a} \tau'(d^*)
\]  

(39)

and can then see that the comparison \(\frac{dRHS(d^*)}{dt_a} \leq \frac{dRHS(d^*)}{dt_u}\) boils down to the same conditions as in (37) and (38) such that \(\frac{dRHS(d^*)}{dt_a} > \frac{dRHS(d^*)}{dt_u}\) if \(t_a + t_u > \tau(d^*) - \nu(d^*)\) and vice versa. This gives rise to the following proposition.

**Proposition 11** *If the market exhibits overprovision (underprovision) competition regulation of the advertiser (consumer) side of the market is more effective than regulation of the consumer (advertiser) side.*
This result is 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 depends crucially on whether the benefit of higher and thus more efficient data provision (non linear) exceeds the increased costs of transportation (linear).\footnote{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 (10)), such that the benefits of decreased transportation costs outweigh the costs from data underprovision.} This trade-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. The resulting second-best level of provided data does not have to be equal to the efficient level of data provision as established in equation (10) because here we additionally weigh transportation costs’ impact on welfare, whereas for the efficient level of data provision as in equation (10) transportation costs were considered as given.

From these results on competition policy we want to draw mainly two 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 Extension

In this chapter we sketch and briefly discuss extensions 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 \( p^u_i \) denote the price a user has to pay to join platform \( i \). User utility is then given by

\[
u_i(x) = v_i + \tilde{d} - \kappa(d_i) - \nu(d_i)a_i - p^u_i - t_c|\tilde{l}_i - x|,
\]

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. Note market shares on both sides of the market now additionally depend on \( p^u_i \) and \( p^u_j \). The resulting profit maximization problem of platform \( i \) is then given by

\[
\max_{p_i, d_i, p^c_i} = a_i \tau(d_i) p_i x_i + p^u_i x_i \quad \forall i \in \{1, 2\}
\]

taking into account profits made from selling access to users. Following the same procedure as in our baseline model we obtain symmetric equilibrium values \( p_i = p_j = \tilde{p}, p^u_i = p^u_j = \tilde{p}^u \) and \( d_i = d_j = \tilde{d} \) where advertiser prices are given by

\[
\tilde{p} = \frac{2 \left( t_a + \nu(\tilde{d}) \right)}{\tau(\tilde{d})},
\]

user prices by

\[
\tilde{p}^u = t_a + t_c + \nu(\tilde{d}) - \tau(\tilde{d}),
\]

while the equilibrium amount of data is given by

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

We immediately see from equations (10) and (44) that \( \tilde{d} = d^o \).

**Proposition 12** 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.
Interpreting advertiser and user prices, as given in equations (43) and (42), we see first that both prices are increasing in advertiser transportation costs $t_a$. It is straightforward that weaker competition on the advertiser side yields higher advertiser prices. The intuition for user prices increasing in $t_a$ is similar to the comparative static effect of $dd^*/dt_a$ as given for (26). Second, user prices increase in $t_u$, i.e. when platform competition for users becomes less intense. Third, it is interesting to note that both prices increase in nuisance $\nu(\bar{d})$.

On the one hand, if users like ads less, platforms will increase ad prices as ads become relatively less important. On the other hand, though, this would deter platform revenues on the advertiser market side, hence they would increase user prices to compensate. Fourth, both prices decrease in targeting effectiveness (or click probability) $\tau(\bar{d})$. Platforms would now rather decrease ad prices, such as to attract some more advertisers, as the extensive margin becomes relatively more important than the intensive margin. Since this would increase the number of ads on a platform, users are accommodated with lower prices such as not to be shied away.

Taking a closer look at equilibrium user prices in (43) we immediately see that negative, positive or zero user prices are possible, depending on parameter values and functional forms.

**Proposition 13** If user prices in the two-sided pricing model are positive (negative), the one-sided pricing constraint would result in data overprovision (underprovision).

**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 9 we know that in the one-sided pricing model too little data is provided, if $t_a + t_u < \tau(d^o) - \nu(d^o)$. 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 9 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 \( \tau(d^o) - \nu(d^o) > \tau(d^*) - \nu(d^*) > t_a + t_u \).

The intuition for this result is that in case of two-sided pricing platforms can extract the efficient amount of data by adequately compensating users. If the market structure is such that the net effect of data collection is large or the competitive environment rather strong, i.e. \( t_a + t_u < \tau(d^o) - \nu(d^o) \), 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 than data than the efficient level. Vice versa, if the net effect of data collection is small or competition rather weak, i.e. \( t_a + t_u > \tau(d^o) - \nu(d^o) \), platforms are not forced to monetize through ads by extracting an inefficiently high amount of data from users as in the one-sided pricing model, but can obtain positive revenue from the user side instead and leave the amount of data at the efficient level.

There are two further conclusions we would like to draw 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 presented in section 3. 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 the platforms maximization problems.

### 7.2 Positive cross-group externalities

In section 3 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 want to show in the following is that the model can in fact be
generalized to have positive cross-group effects in both directions. Consider the following modification of the users’ utility function:

\[ u_i(x) = u - \kappa(d_i) + \rho(d_i)a_i - t_u|l_i - x|. \tag{45} \]

The function \( \rho(d) \) represents the relevance from a user’s point of view of seeing \( a_i \) offers, where \( \rho'(d) > 0 \) and \( \rho''(d) < 0 \). However, \( \rho(d) \) can now be entirely negative, positive or might even switch signs. The first case is discussed in depth in section 3, 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 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.

\[(a) \quad t_u > |\rho(0)|, \rho(d) < tu \]
\[(b) \quad t_a < \tau(0) \]

The upper bound on \(t_a\) is then given by \(\bar{t}_a := \frac{t_u \rho(0) \tau(0)}{3t_u - \rho(0)}\). 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.\(^{17}\) Given that both assumptions are satisfied, the analysis is analogous to our main model and all major results still hold.

### 7.3 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

---

\(^{17}\) In the following we sketch a set of conditions under which both assumptions would be satisfied. Note assumption 4 specifies a lower bound \(t_a > \frac{t_u}{3} \) with \(\frac{t_u}{3} \rho(\cdot) \tau(\cdot)\). 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 \(\lim_{d \to \infty} t_a < t_a\) we can always find intermediate values of \(t_a\) satisfying both conditions. For this to be the case it is necessary that \(\lim_{d \to \infty} t_a < \tau(0)\) and that \(\rho(\cdot)\) is small if positive.
constraint of the indifferent advertiser is binding $p : \pi_i \left( \frac{1}{2} \right) \geq 0$ which yields

$$p = 1 - \frac{t_a}{\tau(d)}. \quad (46)$$

Plugging the collusive price $p$ into the platforms’ profit functions (3) we obtain

$$\Pi_i = \frac{1}{4} (\tau(d) - t_a) \quad (47)$$

and immediately see that profits are increasing in $d$ up to the point where the participation constraint of the indifferent user binds $d : u_i \left( \frac{1}{2} \right) \geq 0$. Since we assumed $u$ 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. This highlights the importance of competition among platforms.

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 substitutes. However, our results suggest that regulation of the advertiser side is more effective if the competitive outcome exhibits overprovision of personal data. Also, a consumer standard would always lead to underprovision of data as consumers do not internalize improvements in the targeting capabilities. Lastly, we showed that two-sided pricing induces platforms to choose the efficient level of data provision and that collusion would always lead to overprovision.

While we think our model provides useful insights we would also like to discuss some limitations. It would be interesting to explore the possibility of endogenous multi-homing on the advertiser side as it would reduce the market power on the advertisers but propagate
the bottleneck property vis-à-vis consumers. Secondly, one could alter the setting on the consumer side and consider heterogeneous consumers, 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 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

\[ \frac{\partial^2 \Pi}{\partial p_t^2} \bigg|_{d^*,p^*} = -\frac{t_u^2 \tau(d^*)^2 (\nu(d^*) + t_u)}{4(t_u - \nu(d^*))^2 (\nu(d^*) \tau(d^*) + t_u t_u)} \]

we immediately see that $\frac{\partial^2 \Pi}{\partial p_t^2} \bigg|_{d^*,p^*} < 0$, 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 $\text{det}(H)|_{d^*,p^*} > 0$.

First, it is helpful to look at the numerator and the denominator of the Hessian separately

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

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

\[ H_{num} = \tau(d^*)^2 \left[ -4t_u^2 (t_u - \tau(d^*)) (\nu(d^*) \tau(d^*) + t_u t_u) (\nu''(d^*)(t_u - \tau(d^*)) + \tau''(d^*) (\nu(d^*) + t_u)) \right. \]
\[ \left. -t_u^2 \nu'(d^*)^2 (t_u - \tau(d^*))^3 - \tau'(d^*)^2 (\nu(d^*) + t_u)^2 (\nu(d^*) (\nu(d^*) (t_u - \tau(d^*)) + 4 t_u \tau(d^*)) + 4 t_u t_u) \right. \]
\[ \left. + 2 t_u \nu(d^*) \nu'(d^*) \tau'(d^*) (t_u - \tau(d^*))^2 (\nu(d^*) + t_u) \right] \]
\[ H_{den} = 64(t_u - \tau(d^*)) (t_u - \nu(d^*))^2 (\nu(d^*) \tau(d^*) + t_u t_u)^2 \]

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

\[ H_{num} = \tau(d^*)^2 [H1_{num} (H2_{num} \nu''(d^*) + H3_{num} \tau''(d^*)) + H4_{num} + H5_{num} + H6_{num}] \]

\[ H1_{num} = -4t_u^2 (t_u - \tau(d^*)) (\nu(d^*) \tau(d^*) + t_u t_u) > 0 \]
\[ H2_{num} = (t_u - \tau(d^*)) < 0 \]
\[ H3_{num} = (\nu(d^*) + t_u) > 0 \]
\[ H4_{num} = -t_u^2 \nu'(d^*)^2 (t_u - \tau(d^*))^3 > 0 \]
\[ H5_{num} = -\tau'(d^*)^2 (\nu(d^*) + t_u)^2 (\nu(d^*) (\nu(d^*) (t_u - \tau(d^*)) + 4 t_u \tau(d^*)) + 4 t_u t_u) \leq 0 \]
\[ H6_{num} = 2 t_u \nu(d^*) \nu'(d^*) \tau'(d^*) (t_u - \tau(d^*))^2 (\nu(d^*) + t_u) < 0 \]
we can see that requiring $H_{num} < 0$ is equivalent to the condition

$$-\frac{1}{H_{1_{num}}} (H_{4_{num}} + H_{5_{num}} + H_{6_{num}}) > H_{2_{num}}\nu''(d^*) + H_{3_{num}}\tau''(d^*)$$

where $LHS \leq 0$ 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 $\nu''(\cdot) > 0$ which is in line with this condition.

### A.2 Proofs for comparative statics

#### Proofs for section 5.2

**Proof.** To see that $d\Pi^P_i/dt_c < 0$, note that

$$\frac{d\Pi^P_i}{dt_a} = -\left[\tau(d^*) - t_a\right] \left[\nu(d^*) - t_u\nu'(d^*) \frac{d\nu'}{d\nu} + \frac{d\tau}{d\nu} \nu(d^*) \right] \left[\frac{\nu''(d^*)}{t_a + \nu(d^*)} \Psi(d^*)\right]$$

$$= \left[\tau(d^*) - t_a\right] \left[(t_u + \nu(d^*)) \nu'(d^*) \tau'(d^*) - \nu'(d^*) \right] \left[\left(\tau(d^*) - t_a\right) [\nu''(d^*) - \nu'(d^*)] - t_c \left(t_u + \nu(d^*)\right) \tau''(d^*)\right]$$

$$\leq 0,$$  \hspace{1cm} (A.1)

where $dd^*/dt_a$ is from equation (23), while $\Psi(d^*)$ is defined in equation (24).

To see that $d\Pi^P_i/d\tau(d^*) > 0$, note that

$$\frac{d\Pi^P_i}{d\tau(d)} = \left[\tau(d^*) - t_a\right] \left[-(t_u + \nu(d^*)) \nu'(d^*) \tau'(d^*) + \nu(d^*) \right] \left[\left(\tau(d^*) - t_a\right) [\nu''(d) + \nu''(d^*)] - t_c \left(t_u + \nu(d^*)\right) \tau''(d^*)\right]$$

$$> 0,$$  \hspace{1cm} (A.2)

where $dd^*/dt_a$ is from equation (23), while $\Psi(d^*)$ is defined in equation (24).
To see that \( \frac{d\pi_i^A}{dt_a} < 0 \), note that

\[
\frac{d\pi_i^A}{dt_a} = \frac{1}{4 [\tau_a + \nu(d^*)]^2} \left\{ -6\tau_a \nu(d^*) - \nu(d^*)^2 \left[ 1 + 2\tau'(d^*) \frac{dd^*}{dt_a} \right] \right.
\]

\[
+ t_c \left[ -\frac{4\nu'(d^*) dd^*}{dt_a} (\tau(d^*) - \tau_a) + t_c \left( -5 + 2\tau'(d^*) \frac{dd^*}{dt_a} \right) \right] \}
\]

\[
= -\frac{1}{4 [\tau_a + \nu(d^*)] \Psi(d^*)} \left\{ -\nu'(d^*) (\tau_a + \nu(d^*)) (\tau(d^*) - \tau_a) \tau'(d^*) + 3 (\tau_a + \nu(d^*))^2 \tau'(d^*)^2 \right.
\]

\[
- (5\tau_a + \nu(d^*)) (\tau(d^*) - \tau_a) \left[ -\nu''(d^*) (\tau(d^*) - \tau_a) + (\tau_a + \nu(d^*)) \tau''(d^*) \right] \}
\]

\[
< 0, \quad \text{(A.3)}
\]

where \( dd^*/dt_a \) is from equation (26), while \( \Psi(d^*) \) is defined in equation (24).

To see that \( \frac{d\pi_i^A}{\nu(d)} < 0 \), note that

\[
\frac{d\pi_i^A}{\nu(d)} = -\frac{[\tau(d^*) - \tau_a]}{2 [\tau_a + \nu(d^*)]^2 \Psi(d^*)} \left\{ (\tau_a + \nu(d^*))^2 \tau'(d^*)^2 \right.
\]

\[
+ 2 [\tau(d^*) - \tau_a] t_c \{ (\tau(d^*) - \tau_a) [\nu''(d) + \nu''(d^*)] - t_c (\tau_a + \nu(d^*)) \tau''(d^*) \} \}
\]

\[
< 0, \quad \text{(A.4)}
\]

References


40


42