Competing for cookies: Platforms' business models in data markets with network effects

Sarit Markovich[*] and Yaron Yehezkel[†]

February 2024

Abstract

We consider platform competition when platforms can either 1) commercialize users' data and in return offer their services for free (data-based business model); 2) protect users' data and charge users for participation (subscription-based model); or 3) offer both options (the hybrid model). We find that competition does not always motivate the incumbent platform to protect users' data. When data has a high public benefit (i.e., users get high benefits from data collected on other users), competition can motivate the incumbent to switch from the data-based to the subscription-based model. Yet, the opposite case occurs when the public benefit of data is small.

JEL Classification: L1

Keywords: platforms with network effects, data commercialization, business models

1 Introduction

Nowadays, platforms manage a massive amount of consumer data. Platforms collect personal consumer information and use it to improve the quality of their service as well as for commercialization purposes, such as selling it to third party vendors or to advertisers. That is, users' data have become an important asset and an essential element of platforms' strategy. This trend is mostly observed in online platforms, where platforms like Google, Facebook, TikTok, and Spotify take advantage of their large stocks of consumer information to offer better products as well as commercialize this data and in return offer their service for free. The reliance and use of users' data has been raising strong concerns about users' privacy.

[*]Kellogg School of Management, Northwestern University (s-markovich@kellogg.northwestern.edu)

[†]Coller School of Management, Tel Aviv University (yehezkel@tauex.tau.ac.il)

In response, regulators have been considering different approaches such as the California Consumer Privacy Act of 2018 (CCPA) which provides consumers with more control over their personal information, or the EU General Data Protection Regulation (GDPR) which provides strict data privacy rules. For example, the GDPR requires that platforms collect and process only as much data as absolutely necessary for the purposes specified and does not allow platforms to discriminate across users that do and do not share their data. Still, some believe that competition would provide strong enough incentives for platforms to choose a business model that preserves users' privacy. Indeed, platforms like Netflix, Apple Health, and Ride with GPS,[1] rely on subscription revenues rather than the commercialization of their users' data. This raises the question: how does competition in data markets affect platforms' choice of business model? Specifically, does it motivate platforms to adopt a business model that does not rely on commercializing users' data? Moreover, under what market conditions does platforms' optimal business model rely on the commercialization of their users' data?

To study these questions, we develop a game with two platforms and users that care about their privacy—i.e., bear a cost if their data is commercialized. Users' disutility from the commercialization of their data differs across users. That is, some users are more sensitive to their privacy than others. Platforms can choose between three business models: (1) data-based; (2) subscription-based; and (3) hybrid. Under the data-based business model, the platform's source of revenue is the commercialization of its users' data. The platform collects data on its users and uses it to improve its service as well as for commercialization purposes, i.e., selling it to third party providers or advertisers. Users that join the platform must share their data with the platform, knowing that the data will be commercialized and the cost this would impose on them. Under the subscription-based business model, users must pay a subscription fee to participate in the platform. The platform still collects users' data but uses it only to improve its service and thus no privacy-cost is imposed on the user. The hybrid business model combines the two first business models. That is, the platform allows users to choose whether they want to join the platform for free and share their data, knowing the data would be commercialized. Alternatively, users can pay the subscription fee, in which case their data will not be commercialized. For example, Google, Facebook, TikTok, and Twitter utilize the data-based business model, while Apple has been an avid advocate of the subscription-based model. Likewise, the social apps True and Mastodon, the messaging app Signal, and the search engine DuckDuckGo, explicitly chose not to commercialize their users' data. True plans on making money by charging users for subscription.[2] Netflix has

---

[1]Ride with GPS is a social route-planning and navigation tool for cyclists.

[2]Mastodon relies on decentralization, Signal on donations, and DuckDuckGo on keywords, rather than targeted, advertising.

only recently switched from the subscription-based model to the hybrid one. In November 2023, Meta launched in Europe a no-ads subscription service. Accordingly, users can choose between a free service by agreeing to have their data tracked and commercialized through advertising, or choose a subscription model which protects their privacy and offers an ad-free experience. This business model is controversial in Europe. A coalition of 28 organizations has called for an investigation of this business model, arguing that Meta essentially asks users to pay for their privacy. In contrast, a spokesperson for Meta cited decision by the Court of Justice of the European Union (CJEU) in July that the hybrid models with a subscription-based option are legitimate means for users to consent to data processing for personalized advertising.[3]

An important feature of our model is that users bear costs when their data is commercialized but not from the mere collection of their data. Moreover, because platforms collect data from all users and use it to improve the quality of their service, users enjoy the benefits of data collected on all users that join the platform, regardless of whether their data is commercialized or not. That is, data collected on one user provides public benefit to all the users on that platform. Following Markovich and Yehezkel (2023), we refer to the benefits associated with the overall data collected by a platform as the public benefit of data. For example, data that a navigation app collects from a driver can benefit other drivers that consider taking the same route. Other relevant examples with high public benefits of data are users that provide their location data on a contact-tracing app benefit others who now know they were in proximity of someone who tested positive for COVID-19;[4] or Fitbit's use of its heart rate data to identify episodes of irregular heart rhythm suggestive of atrial fibrillation (AFib), the most common form of heart rhythm irregularity. Fitbit intends to use this information to alert users about an irregular heart rhythm so that notified individuals would connect with a doctor. This is in contrast to apps like Ride with GPS where the public benefit of data–the ability to see others' routes–is much lower. We show the that the magnitude of the public benefit of data has a determining effect on the platforms' equilibrium business models. Moreover, the public benefit of data affects the way competition changes platforms' optimal choice of business model.

In order to capture the advantage that a large, dominant platform may have, we assume a two-stage game with an incumbent and an entrant, where the incumbent enjoys a focality advantage. That is, users believe that the incumbent would be the dominant platform in the market. Users can join one of the platforms or stay out. Since we are interested in isolating

---

[3]See CPI, February 18, 2024. Available at: https://www.pymnts.com/cpi_posts/privacy-advocates-urge-european-regulators-to-oppose-metas-no-ads-subscription-model/

[4]Contact tracing apps use one's phone, or other mobile device, to track and alert individual if they'd crossed paths with someone who within a certain window of time tested positive to COVID-19.

the effect of competition on platforms' choice of business model, we assume that the base quality of the entrant and the incumbent are the same.

We find that the effect of competition on platforms' optimal business model depends on the interaction between the public and commercial benefit of data. Starting with the case where platforms can only choose between a data-based and a subscription-based business model. We show that if the commercial benefit of data is not too high or too low and the public benefit of data is high, competition incentivizes a monopolistic incumbent to move away from a data-based model and promote privacy by choosing the subscription-based. If, however, the public benefit of data is low, then the opposite behavior happens, and competition incentivizes a monopolistic incumbent to go with a data-based model rather than a subscription-based one. That is, competition does not necessarily promote a more privacy sensitive market.

Our second main result concerns with the hybrid business model. The GDPR does not allow platforms to discriminate across users that share their data and those who do not share. That is, the GDPR bans platforms from offering the hybrid business model we discussed above. One may wonder whether the ability to offer a hybrid model promotes a more or less privacy-sensitive market. We find that, as long as the commercial benefit of data is small, the ability to offer a hybrid model has no effect on the incumbent's choice of business model. Once the commercial value of data is high enough, the effect of the ability to offer a hybrid business model on platforms' business model choice largely depends on the strength of the public benefit of data. In particular, when the public benefit of data is intermediate (high), the incumbent chooses the hybrid model over the data-based (subscription-based) one. In both these cases, the hybrid model is preferable because it allows the incumbent to enjoy "both worlds" and dominate the market. Interestingly, when the public benefit of data is low, the availability of the hybrid business model prompts the incumbent to choose the subscription-based business model over the data-based it chooses when hybrid is not an option. Here, it is the threat of entry and the entrant's ability to offer a hybrid business model that affects and determines the incumbent's choice of business model. Specifically, the incumbent knows that if it chooses the data-based model the entrant would choose the hybrid one and monopolize the market. Consequently, the incumbent chooses a more privacy-sensitive business model and shares the market with the entrant which chooses in response the data-based business model.

Finally, we also look at how competition affects the selected business model when platform can adopt the hybrid model. As in the case where platforms can only adopt the data-based or the subscription-based regimes, we find that competition affects the incumbent's business model (in comparison with the case in which the incumbent is a monopoly) when

the commercial benefit of data is intermediate. Here, if the public benefit of data is high, competition motivates a monopolistic incumbent to switch from a subscription-based model to the hybrid model. Indeed, Netflix choose to offer a hybrid business model followed the increased competition in the streaming market. In contrast, when the public benefit of data is small, competition motivates the incumbent to switch from the hybrid model to the subscription-based model. That is, competition may decrease the attractiveness of the hybrid business model, in which case the incumbent chooses to focus on the privacy-sensitive users. We further show that the hybrid model may implement the welfare maximizing outcome, but may result in over-commercialization of data if the public benefit of data is high, and under-commercialization of data otherwise.

Our paper extends the literature on platform competition to the case where the market exhibits network effects and platforms' business models considers the protection of users' privacy. The paper closest to ours is Casadesus-Masanell and Hervas-Drane (2015) who study a competitive market where firms compete in prices and qualities, which can be interpreted 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. They also show that low privacy firms tend to subsidize consumers, while high privacy firms charge positive prices. The main contribution of our paper is the introduction of network effects, in the form of the public benefit of data. We show that this public benefit has a qualitative effect on how an incumbent platform responds to competition.

The economic literature on competing platforms (see Jullien, Pavan and Rysman, 2021, for a review of the literature) extends the work of Katz and Shapiro (1985) on competition with network effects, where the size of the network creates additional value to the customers. Jullien (2011), Hałaburda and Yehezkel (2013; 2016; 2019) consider platform competition and coordination in the context of a static game. Hagiu (2006) considers sequential competition on two sides of a market. Hałaburda et al. (2020) and Biglaiser and Crémer (2020) consider dynamic competition. Much of this literature focuses on the coordination problem and the role pricing plays in overcoming this problem by using a divide-and-conquer strategy where platforms compete in subsidizing one set of users in order to attract another set. Our paper is closer to Markovich and Yehezkel (2022) who study platform competition with user-group. Similar to their result where platforms compete on attracting the group because the group determines which platform wins the non-group users, in the current paper, platform compete on attracting the users who are not data-sensitive because the data collected on the increases the public benefit the platform can offer to attract the data-sensitive users.

Our paper is also related to the literature on privacy and network externalities. Most of this existing literature o focuses on the negative externalities associated with users shar-

ing their data where one user's data can help platforms learn and predict the behavior of other users who do not share their data (Fairfield and Engel, 2015; Choi, Jeon, and Kim, 2019; Acemoglu et al., 2019; Bergemann, Bonatti, and Gan, 2022; Liang and Madsen, 2019). Following Markovich and Yehezkel (2023), our paper recognizes and focuses on the positive externalities—e.g., users that share data help the platform improve the quality of its product and offer higher value to other users. Fainmesser et al. (2022) study how a monopolistic platform's revenue model affects its data policy in terms of data collection and data protection. Considering the net value of network externalities (positive minus negative), they find that relative to the socially desired data strategy, the platform may over- or under-collect users' data and may over- or under-protect it. The authors then show that the inefficiency in data collection can be corrected with taxes or fines imposed on the firms. We add to this literature by focusing on competition and its effect on platforms' business models in terms of commercializing data or charging users for using the platform's. O'Brien and Smith (2014) study a model where sellers can commit to privacy policies and consumers have heterogeneous – negative or positive – preferences over privacy. They find that under perfect competition, firms make the socially optimal decision. Furthermore, a positive and sufficiently large correlation between consumers' valuations for the product and privacy is a necessary condition for the under-supply of privacy by firms. Assuming a two-stage game where data accumulated in the first period can be used to customize products in the second stage, Ke and Sudhir (2022) find that in a perfectly competitive market, whether privacy rights lower or increase profits depends on the expected privacy breach costs. Our paper considers imperfect competition between an incumbent and an entrant platforms. We show how the strategic effect of competition and the threat of entry shape the the incumbent's and the entrant's business models. Similar to our paper, Hagiu and Wright (2023) study competition between an incumbent and entrant platform that collect data on their users. The focus of their analysis, however, is on data-enabled learning across- and within-users and on how a platform's competitive advantage is affected by the shape of the learning function.

Our paper is also related to the growing empirical literature studying the impact of the GDPR. Utilizing data from an online travel intermediary, Aridor et al. (2022) find that the GDPR has resulted in an immediate drop in the total number of advertisements clicked and a corresponding immediate decline in revenue. The remaining set of consumers, however, are higher value consumers to the advertisers, compared with the pre-GDPR set of consumers. Focusing on market concentration, Johnson et al. (2023) find that GDPR increased market concentration among technology vendors where the relative market shares of the largest firms—particularly, Google and Facebook—increase post-GDPR. Using data on apps at the Google Play Store Janssen et al. (2022) show that GDPR induced the exit and reduced

6

entry of new apps by half, resulting in an overall reduced consumer surplus. We add to this literature by analyzing the effect of banning firms from the ability to using a hybrid business model which price discriminates between users that share their data and those who do not share their data for commercialization.

2 The Model

Consider two competing platforms, an incumbent, I, and an entrant, E, and a mass 1 of users. Each platform can collect data from users and utilize the data for two benefits. The first benefit is enhancing services to other users. We refer to this benefit as the users' "public benefit" of data and we denote it by β. Secondly, the platform can commercialize the data by selling it to advertisers or other platforms. We refer to it as the platform's "commercial benefit" and we denote it by α. Users incur disutility when their data is commercialized, denoted by k. A user's k's utility from joining platform $i = I, E$ is:

$$U_{ki} = v + \beta n_i - C_i k - p_i, \tag{1}$$

where v is the base benefit from joining a platform,[5] n_i is the number of users that join platform i, $C_i = \{0, 1\}$ is the platform decision on whether not to commercialize the user's data ($C_i = 0$) or to commercialize ($C_i = 1$), in which case the user incurs a costs k. Finally, p_i is the platform's price. Suppose that users differ in their costs for commercializing their data: some users are more sensitive to their privacy than others. Hence, we assume that k is uniformly distributed on the interval $[0, 1]$. We focus on the interesting case where the market is not fully covered, when the platform commercializes users' data and thus restrict the parameter space to: $v < 1$ and $0 < \beta < 1 - v < 1/2$. Moreover, this parameter space rules out corner solutions where users gain negative utility under price competition.

Each platform can choose between three business models: data-based that we denote by D, subscription-based that we denote by S and a hybrid model, denoted by H. In the data-based business model, $C_i = 1$: the use of the platform is free and its source of revenues is from collecting and commercializing users' data. In this case, the platform's profit is $\pi_i(D, B_j) = \alpha n_i(D, B_j)$, where $n_i(D, B_j)$ is the number of users that join it given that platform j adopts business model $B_j = D, S, H$; recall that $\alpha > 0$ is the data's commercial benefit to the platform. Under the "subscription based" business model, the platform commits not to commercialize users' data ($C_i = 0$) and instead charges users for participation and

---

[5]Our analysis focuses on the effect of platforms' choice of business model on competition. In order to isolate this effect, we assume that both platforms offer the same base benefit.

earns $\pi_i(S, B_j) = p_i n_i(D, B_j)$. The third, hybrid business model is a combination of the two: the platform allows users to choose between a subscription plan in which it commits not to commercialize the user's data and a free plan where it makes no such commitment and hence commercializes users' data. The platform's profit is $\pi_i(H, B_j) = \alpha n_{iD}(H, B_j) + p_i n_{iS}(H, B_j)$, where $n_{iD}(H, B_j)$ and $n_{iS}(H, B_j)$ are the number of users that join the free and subscription plan, respectively.

The timing is as follows. In the first stage, the incumbent chooses its business model: $B_I = D, S, H$. In the second stage, the entrant chooses its business model $B_E = D, S, H$. Then, in the third stage, the two platforms compete on users. In sections 3-4, we assume that, due to regulation, platforms cannot adopt the hybrid model. Then, in Section 5, we show how allowing platforms to adopt the hybrid model affects the results.

As is usually the case in platform competition with network effects, in the third stage of the game there can be multiple equilibria, because each user's decision depends on the beliefs regarding the decisions of other users. We assume that the incumbent has a "focal" position in that whenever possible, users expect other users to join the incumbent. We elaborate on these beliefs in Section 4.

3   The monopoly benchmark

Before solving for competition, we start with the benchmark case in which the incumbent is a monopoly. The incumbent has two options. First, if the incumbent chooses the data-based business model, it announces its intention to commercialize users' data. Users will join the incumbent as long as:

$$v + \beta n_I - k \geq 0 \iff n_I(D) = \frac{v}{1 - \beta}. \tag{2}$$

Because by assumption $v < 1 - \beta$, not all users join the platform: data-sensitive users, users with a high k, prefer to stay out. Yet, as the public benefit of data increases, more users join the platform in order to enjoy the public benefit of data generated by other users. The incumbent earns $\pi_I(D) = \frac{\alpha v}{1 - \beta}$. Suppose now that the incumbent adopts the subscription-base business model. Given $p_I > 0$ and $C_I = 0$, because the incumbent benefits from a focal position and users expect other users to join it, the incumbent can attract all users if $p_I \leq v + \beta$. Hence, the incumbent charges $p_I(S) = v + \beta$ and earns $\pi_I(S) = v + \beta$. Comparing $\pi_I(D)$ and $\pi_I(S)$ yields the following result:

Corollary 1. *A monopolistic incumbent chooses the data-based business model ($\pi_I(D) >$*

$\pi_I(S)$) iff:

$$\alpha \geq \alpha_{D,S}^M = \frac{(1-\beta)(\beta+v)}{v},\tag{3}$$

where $\alpha_{D,S}^M$ is an inverse U-shape function of the public benefit of data, β.

Intuitively, the incumbent adopts the data-based business model if the commercial benefit is above some threshold. Yet, this threshold is non-monotonic in the public benefit of data: the incumbent's incentive to adopt the data-based business model is first decreasing in β and then decreasing in it. The intuition for this result is that under the subscription-base business model, an increase in β results in a one-to-one linear increase in users' (and hence the platform's) value. This is not the case in the data-base business model, where an increase in β results in a concave increase in $\pi_I(D)$ due to the following two effects: (i) for a given $n_I(D)$, a higher public benefit of data translates to a higher utility for each user, hence more users join and share their data; (ii) the increase in $n_I(D)$ further increases the public benefit, which in turn increases again the number of users that join and share their data. The combined effect creates a convex relationship between β and $\pi_I(D)$, where the effect of β is stronger as β increases. Because this non-linear effect of β on the profitability of the data-base business model, this business model is attractive for high and low values of β, while the subscription-base business model is more attractive for intermediate values of β.

4   Competition

Suppose that the incumbent competes with an entrant. The incumbent chooses its business model first, $B_I = D, S$, followed by the entrant, $B_E = D, S$, and then the two platforms compete on consumers. We solve the game backwards, and start by solving each of the 4 market configurations: $(B_I, B_E) = \{(D,S), (D,D), (S,S), (S,D)\}$.

Platforms adopt different business models

Suppose that the incumbent chooses the data-based business model and the entrant chooses the subscription-based model: $(B_I, B_E) = (D, S)$. Hence, $C_I = 1$ and $C_E = 0$.

We first note that there is no equilibrium in which the incumbent dominates the market, which makes focality irrelevant in this business model configuration. If such an equilibrium were to exist, $p_E = 0$ and all users join the incumbent. Yet, even when all users join the incumbent, the utility of the most data-sensitive user with $k = 1$ from joining the incumbent is $1 \times \beta + v - 1$, which is lower than the utility v that the user can gain by joining the entrant, because of our assumption that $\beta < 1 - v$. We, therefore, solve for an equilibrium in which

the entrant gains a positive market share. Intuitively, adopting different business models creates differentiation, which enables the entrant to gain positive market share despite the incumbent's focality advantage.

In equilibrium, given that n_E users join the entrant and $n_I = 1 - n_E$ users join the incumbent, there is a user, $\widehat{k}$, who is indifferent between joining the incumbent or the entrant. This user solves:

$$\beta(1 - n_E) + v - \widehat{k} = \beta n_E + v - p_E.$$

When there is an internal solution to $\widehat{k}$ (i.e., $0 < \widehat{k} < 1$), users with $k \in [0, \widehat{k}]$ join the incumbent because they are not sensitive to their privacy and therefore prefer a free service, even if the platform commercializes their data. In contrast, data-sensitive users with $k \in [\widehat{k}, 1]$ prefer the platform that charges a membership fee in order to protect their privacy. Hence, the demand function facing the entrant is:

$$n_E(p_E) = \frac{1 - \beta - p_E}{1 - 2\beta}. \tag{4}$$

Because $\beta < 1/2$, the denominator in (4) is positive. Yet, notice that β has two conflicting effects on the demand facing the entrant. To see how, the inverse demand function of (4) is $p_E(n_E) = 1 - \beta - (1 - 2\beta)n_E$, which rotates counterclockwise around $n_E = 1/2$ as β increases, such that the demand increases with β if $n_E > 1/2$ and decreases with β otherwise. The intuition for this feature of the demand function is that when $n_E > 1/2$, the entrant, who does not commercialize users' data, serves more users than the incumbent and thus also collects more data. Hence, as the public benefit of data increases, the entrant's demand increases. The opposite case occurs when $n_E < 1/2$.

The entrant sets p_E to maximize $\pi_E(p_E) = p_E n_E(p_E)$:

$$p_E(D, S) = \begin{cases} \frac{1-\beta}{2}, & \text{if } \beta \leq \frac{1}{3}, \\ \beta, & \text{if } \beta > \frac{1}{3}, \end{cases} \qquad n_E(D, S) = \begin{cases} \frac{1-\beta}{2(1-2\beta)}, & \text{if } \beta \leq \frac{1}{3}, \\ 1, & \text{if } \beta > \frac{1}{3}. \end{cases}$$

As a technical note, recall that the constraint $\beta < 1 - v$ implies that the second row in $p_E(D, S)$ and $n_E(D, S)$ are relevant only when $v < \frac{2}{3}$.

The entrant's price decreases in β while the entrant's market share increases in it.[6] Intuitively, at $\beta = 0$, the two platforms equally share the market. As β increases, the entrant's price decreases while its market share increases, because the entrant can better exploit the increase in the public benefit for enhancing its demand. Moreover, because the entrant does

---

[6]We verified that the utility of the indifferent user is always positive because $v > 1/2$, hence all users gain positive utility from joining a platform.

not commercialize users' data, the entrant can fully dominate the market if β is sufficiently high. The profits of the two platforms in the $(B_I, B_E) = (D, S)$ business model configuration are $\pi_E(D, S) = p_E n_E(p_E)$ and $\pi_I(D, S) = \alpha(1 - n_E(p_E))$, or:

$$\pi_E(D, S) = \begin{cases} \frac{(1-\beta)^2}{4(1-2\beta)}, & \text{if } \beta \leq \frac{1}{3}, \\ \beta, & \text{if } \beta > \frac{1}{3}, \end{cases} \qquad \pi_I(D, S) = \begin{cases} \frac{\alpha(1-3\beta)}{2(1-2\beta)}, & \text{if } \beta \leq \frac{1}{3}, \\ 0, & \text{if } \beta > \frac{1}{3}. \end{cases} \qquad (5)$$

The following corollary summarizes the features of the (D, S) market configuration.

Corollary 2. *Suppose that the incumbent adopts a data-based business model and the entrant adopts a subscription-based one. Then, the entrant's price decreases in the public benefit of data yet its market share increase with it. Moreover, if the public benefit of data is sufficiently high, the entrant dominates the market .*

Finally, the opposite business-model configuration in which the incumbent chooses the subscription-based business model and the entrant chooses the data-based business model is symmetric: $\pi_E(S, D) = \pi_I(D, S)$ and $\pi_I(S, D) = \pi_E(D, S)$.

Both platforms adopt the data-based business model

Suppose now that $(B_I, B_E) = (D, D)$. Hence, $C_I = C_E = 1$. As both platforms adopt the same business model, there are two equilibria. In both equilibria, all users who join a platform make the same decision: they either all join the incumbent or they all join the entrant. In particular, in both equilibria, $n_i = \frac{v}{1-\beta}$ users join platform i and the remaining users (which are the data-sensitive users) stay out of both platforms. This is an equilibrium because the user with $k = n_i$ is indifferent between joining a platform or staying out given the expectation that $n_i = \frac{v}{1-\beta}$ users join platform i, and because when all users expect that $n_i = \frac{v}{1-\beta}$ and $n_j = 0$ $(j \neq i)$, all users who join a platform prefer to join platform i.

To solve the problem of multiple equilibria, we follow the literature on platform competition (Caillaud and Jullien (2001; 2003), Hałaburda and Yehezkel (2016)) and assume that the incumbent if "focal": when both equilibria are possible, users expect all other users to join the incumbent. Intuitively, because the incumbent already exists in the market, costumers' inertia creates the expectations that users will continue to prefer the incumbent over he entrant. Given that the incumbent is focal, it therefore dominates the market. The following Corollary summarizes these results:

Corollary 3. *Suppose that both platforms adopt the data-based business model. Then, the incumbent dominates the market, serves $n_i = \frac{v}{1-\beta}$ users and earns $\pi_I(D, D) = \frac{\alpha v}{1-\beta}$ while the entrant earns $\pi_E(D, D) = 0$.*

11

Both platforms adopt the subscription-based business model

Suppose that $(B_I, B_E) = (S, S)$; i.e., $C_I = C_E = 0$. Given the platforms' prices, p_I and p_E, there is an equilibrium in which all users join the incumbent if: $\beta + v - p_I \geq v - p_E$ or $\beta \geq p_I - p_E$. Likewise, there is an equilibrium in which all users join the entrant if $\beta + v - p_E \geq v - p_I$ or $p_I - p_E \geq -\beta$. As the two conditions overlap, for $\beta \geq p_I - p_E \geq -\beta$ there are two equilibria in which either the incumbent or the entrant wins the market. Given our focality assumption, in this case all users play the equilibrium in which they join the incumbent. Hence, the equilibrium prices are $p_E = 0$, $p_I = \beta$ and the incumbent dominates the market. The following Corollary summarizes these results:

Corollary 4. *Suppose that both platforms adopt the subscription-based business model. Then, in equilibrium, $p_E = 0$, $p_I = \beta$ and the incumbent wins the market. Profits are $\pi_I(S, S) = \beta$ and $\pi_E(S, S) = 0$.*

Equilibrium business models

Consider now the first and second stage in which the incumbent chooses its business model followed by the choice of the entrant. Consider first the case where $\beta < \frac{1}{3}$ (or, when $v > \frac{2}{3}$, consider the case where $\beta < 1 - v$). Because the entrant loses the market if it chooses the same business model as the incumbent's, the entrant always chooses the opposite business model than the incumbent. Taking that into account, the incumbent adopts the data-based business model if and only if:

$$\pi_I(D, S) > \pi_I(S, D) \iff \alpha > \alpha^C_{D,S} = \frac{(1 - \beta)^2}{2(1 - 3\beta)}. \tag{6}$$

The following proposition compares between this threshold value of α and the threshold value under monopoly (proofs of all propositions are in the Appendix):

Proposition 1. *Suppose that $\beta < \frac{1}{3}$. If the public benefit of data is low, the incumbent has a stronger incentive to adopt the data-based business model under competition than under monopoly. Yet, if the public benefit of data is high, the incumbent has a weaker incentive to adopt the data-based business model under competition than under monopoly. That is, there is a threshold, $\overline{\beta}$, such that $\alpha^C_{D,S} < \alpha^M_{D,S}$ if $\beta < \overline{\beta}$ and $\alpha^C_{D,S} > \alpha^M_{D,S}$ otherwise.*[7]

Figure 1 illustrates the results of Proposition 1.[8] The figure shows that when the public benefit of data is low (i.e., $\beta < \overline{\beta}$), then for intermediate commercial benefit ($\alpha^C_{D,S} < \alpha <$

---

[7]The proof further shows that when $v < \frac{2}{3}$, $\overline{\beta} < \frac{1}{3}$. When $v > \frac{2}{3}$, $\overline{\beta} < 1 - v$ when $v < 3 - \sqrt{5} \cong 0.764$.
[8]The figure assumes that $v < \frac{2}{3}$, such that the binding constraint on β is $\beta \leq \min\left\{\frac{1}{3}, 1 - v\right\} = \frac{1}{3}$.
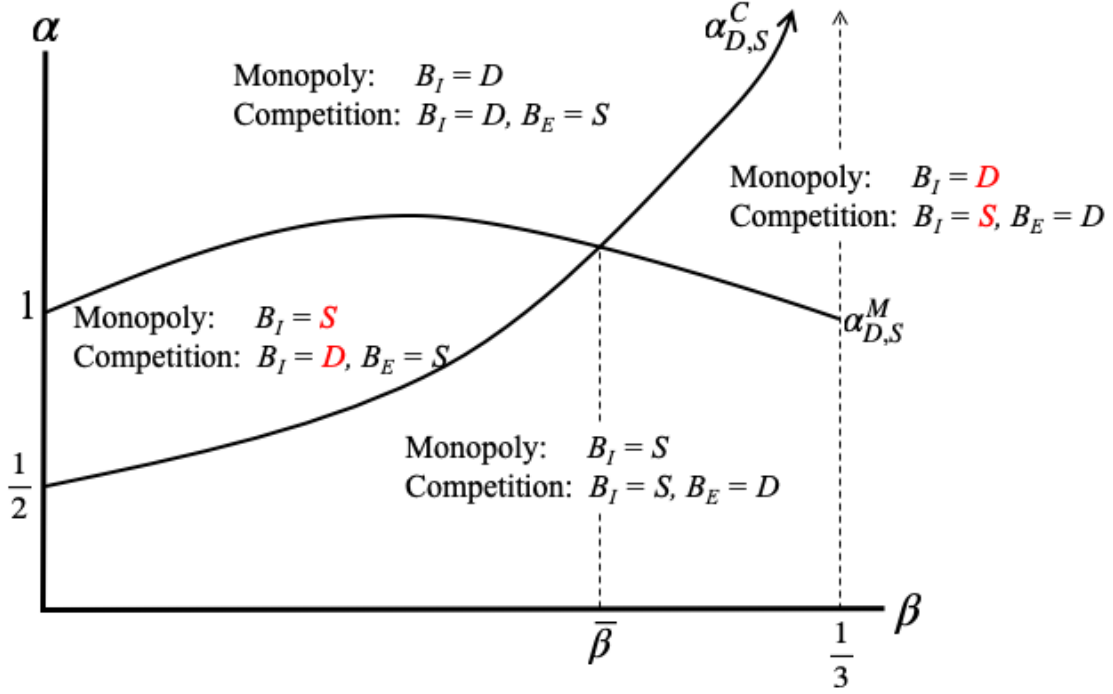
Figure 1: $\alpha_{D,S}^M$ and $\alpha_{D,S}^C$ as a function of β and the equilibrium business models (when $v < \frac{2}{3}$)

$\alpha_{D,S}^M$), competition incentivizes a monopolistic incumbent to switch from the subscription-based business model to the data-based one. Yet, when the public benefit of data is high (i.e., $\beta > \overline{\beta}$), then for intermediate commercial benefit ($\alpha_{D,S}^M < \alpha < \alpha_{D,S}^C$), competition induces a monopolistic incumbent to change its business model from data-based to subscription-based. That is, competition does not necessarily promotes a more privacy sensitive market.

To see the intuition for this result, it is useful to identify two conflicting effects that competition has on the optimal business model:

Proposition 2. *Compared to the monopoly benchmark, competition reduces the incumbent's market share in the data-based business model, but reduces the incumbent's price in the subscription-based business model. That is, $n_I(D) > n_I(D, S)$ and $p_I(S) > p_I(S, D)$.*

Proposition 2 states that on the one hand, if the incumbent chooses the data-based business model, competition reduces its market share because in this case the entrant adopts the subscription-based model and as a result attracts users with intermediate k that under monopoly would have given up on their privacy and joined the incumbent. Due to this effect, competition motivates the incumbent to switch from the data-based to the subscription-based model. On the other hand, as the proposition notes, if the incumbent adopts the subscription-based model, the incumbent charges a lower price under competition than under monopoly. As shown in Proposition 1, the magnitude of these two effects depends on the data's public benefit. Intuitively, as β increases, more users attract more users. Hence, the first effect

becomes stronger. Once β is above some threshold, the first effect dominates and competition incentivizes the incumbent to switch from the data-based to the subscription-based model. The opposite case occurs when β is below the threshold.

Finally, when the public benefit is very high ($\beta > \frac{1}{3}$), in equilibrium, the incumbent always chooses the subscription-based model because otherwise, the entrant can dominate the market. Here, a monopolistic incumbent would rather adopt the data-based model only if α is high enough. Therefore, competition can motivate the incumbent to switch from the data-based model to the subscription-based model.

5   The hybrid business model

Up until now, we assumed that platforms can choose only one of the extreme business models–only data-based or only subscription-based. A third potential business model that platforms can adopt is a hybrid model (denote by H), in which the platform offers both a subscription plan and a free plan. The platform commits not to commercialize data of users who join the subscription plan, but makes no such commitment for users who join the free plan, hence commercializes their data. Users can freely choose whether to join the subscription plan or the free plan, or not join the platform.

In this section we assume that platforms can adopt the hybrid model, and compare the profitability of the three models. The main conclusion is that when the data's commercial benefit is small or when the public benefit is small, the incumbent prefers the subscription-based model and the entrant responds by adopting the data-based model. Yet, when the commercial benefit of data and the public benefit are both high, the incumbent adopts the hybrid model and monopolizes the market. That is, the hybrid business-model allows the incumbent to deter entry.

The platforms' ability to adopt the hybrid model open the possibility for additional market configurations: $(B_I, B_E) = (H, B_E)$ and $(B_I, B_E) = (B_I, H)$. We first solve the outcome when the incumbent adopts the hybrid model, and then the cases where the entrant responds to the incumbent's business model by adopting the hybrid model.

The incumbent adopts the hybrid model

Suppose that the incumbent adopts the hybrid model. The incumbent announces that users can either join for free, conditional on giving their consent to have their data commercialized, or pay a price, p_I, and have their data protected.

Consider an equilibrium in which the incumbent dominates the market. The entrant's

optimal response is to adopt the subscription-based model and offer it for free. Doing so provides users with the highest alternative utility to the utility from joining the incumbent. As the incumbent benefits from a focal position, users therefore expect that all other users join the incumbent, and users' utility from joining the entrant is $0 \times \beta + v - p_E = v$.

Turning to the incumbent, given the price of the subscription plan, p_I, users who join the incumbent choose the subscription plan if $\beta + v - p_I \geq \beta + v - k$, or $k \geq p_I$. Users with $k \in [0, p_I]$ join the free plan and the incumbent commercializes their data and earns αp_I. Users with $k \in [p_I, 1]$ join the subscription plan, pay p_I, and the incumbent earns from these users $(1 - p_I)p_I$. Hence, the incumbent's maximization problem is to choose p_I that maximizes:

$$\max_{p_I} \pi_I(p_I|(H, S)) = \alpha p_I + (1 - p_I)p_I, \tag{7}$$

$$\text{s.t.} \quad \beta + v - p_I \geq v \text{ and } p_I \leq 1.$$

The first constraint requires that the user who is indifferent between joining the incumbent's data-based plan and the subscription-based plan prefers these options over joining the entrant's subscription plan for free. The second constraint requires that there is an internal solution to the indifferent user. The unconstraint solution is $p_I = (1 + \alpha)/2$. Notice that users who join the subscription plan gain the utility $\beta + v - p_I = \beta + v - \frac{1}{2}(1 + \alpha) < v$, where the inequality follows because $\beta < \frac{1}{2}$. Hence, the maximization problem has a corner solution in which the binding constraint is: $\beta + v - p_I > v$, or $p_I = \beta$. The incumbent earns from the hybrid model:

$$\pi_I(H, S) = (1 + \alpha - \beta)\beta,$$

and $\pi_E(H, S) = 0$. The following corollaries summarize the result.

Corollary 5. *If the incumbent adopts the hybrid business-model, and the entrant can choose between $B_E = \{D, S, H\}$, then the incumbent dominates the market, charges $p_I = \beta$ and earns $\pi_I(H, S) = (1 + \alpha - \beta)\beta$. The incumbent serves all users and commercializes the data of users with $k < \beta$.*

The incumbent adopts the data-based model

Suppose now that the incumbent adopts the data-based model. From Section 4, it is clear that the entrant will respond by either adopting the subscription-based model (as we solved in Section 4) or by adopting the hybrid model. For brevity, we analyze these two cases in the appendix and state here the following result:

Proposition 3. *Suppose that the incumbent adopts the data-based model ($B_I = D$), and the*

entrant can choose between $B_E = \{D, S, H\}$. Then, there is a threshold,

$$\overline{\alpha} = \begin{cases} \frac{1-\beta}{\sqrt{1-2\beta}} - 1, & \text{if } \beta < \frac{1}{3}, \\ 2\sqrt{\beta} - 1, & \text{if } \beta > \frac{1}{3}, \end{cases}$$

such that:

(*i*)   *For $0 < \alpha < \overline{\alpha}$, the entrant adopts the subscription-based model, $B_E = S$. The outcome is identical to the outcome in Section 4 when platforms adopt different business models;*

(*ii*)   *For $\overline{\alpha} < \alpha$, the entrant adopts the hybrid model, $B_E = H$, and dominates the market.*

Intuitively, the entrant never responds to the incumbent's data-base model by adopting the data-base model because then it loses the market. If data has low commercial value (α is small), then the entrant prefers the subscription based model which does not rely on commercializing user' data. Otherwise, the entrant can use the hybrid model to attract users with high disutility from data commercialization with the subscription plan, and use the public benefit of their data to attract the less data-sensitive users with a free plan. This way, the entrant can dominate the market.

The incumbent adopts the subscription-based model

Suppose now that the incumbent adopts the subscription-based model. Recall that if the entrant also adopts the subscription-based model, the incumbent wins the market due to its focal position. This logic follows to the case where the entrant adopts the hybrid model. Even if the entrant charges $p_E = 0$ and the incumbent charges $p_I = \beta$, there is an equilibrium in which all users join the incumbent and do not share data, because $\beta \times 1 + v - p_I \geq \beta \times 0 + v - p_E$. The following corollary summarizes this result:

Corollary 6. *Suppose that the incumbent adopts the subscription-based model ($B_I = S$), and the entrant can choose between $B_E = \{D, S, H\}$. Then, the entrant adopts the data-based model and the platforms' profits are the same as in Section 4.*

Equilibrium business model

We can now turn to solving the equilibrium business models when both platforms can adopt $B_i \in \{D, S, H\}$. We start with the case in which $\beta < \frac{1}{3}$. The following proposition shows when it is optimal for the incumbent to adopts the hybrid model:

Proposition 4. *Suppose that $\beta < \frac{1}{3}$. Then, the incumbent adopts the hybrid model when the public benefit and the commercial benefits are high, and the subscription-based model otherwise. That is, there is a threshold, $\alpha_{H,S}^C$, where*

$$\alpha_{H,S}^C = \frac{(1-\beta)(1-5\beta+8\beta^2)}{4\beta(1-2\beta)}, \tag{8}$$

such that when $\alpha > \alpha_{H,S}^C$, the incumbent adopts the hybrid model and dominates the market. When $\alpha < \alpha_{H,S}^C$, the incumbent adopts the data-based model while the entrant adopts the subscription based model and the two platforms share the market. Moreover, $\alpha_{H,S}^C$ is decreasing with β.

The hybrid model is more profitable as the commercial benefit of data is high, as this increases the revenues from the data plan. Likewise, the hybrid model is more profitable as the public benefit of data is high, because this increases the revenues from the subscription plan. Taking together, the hybrid model is more profitable for high values of α and β.
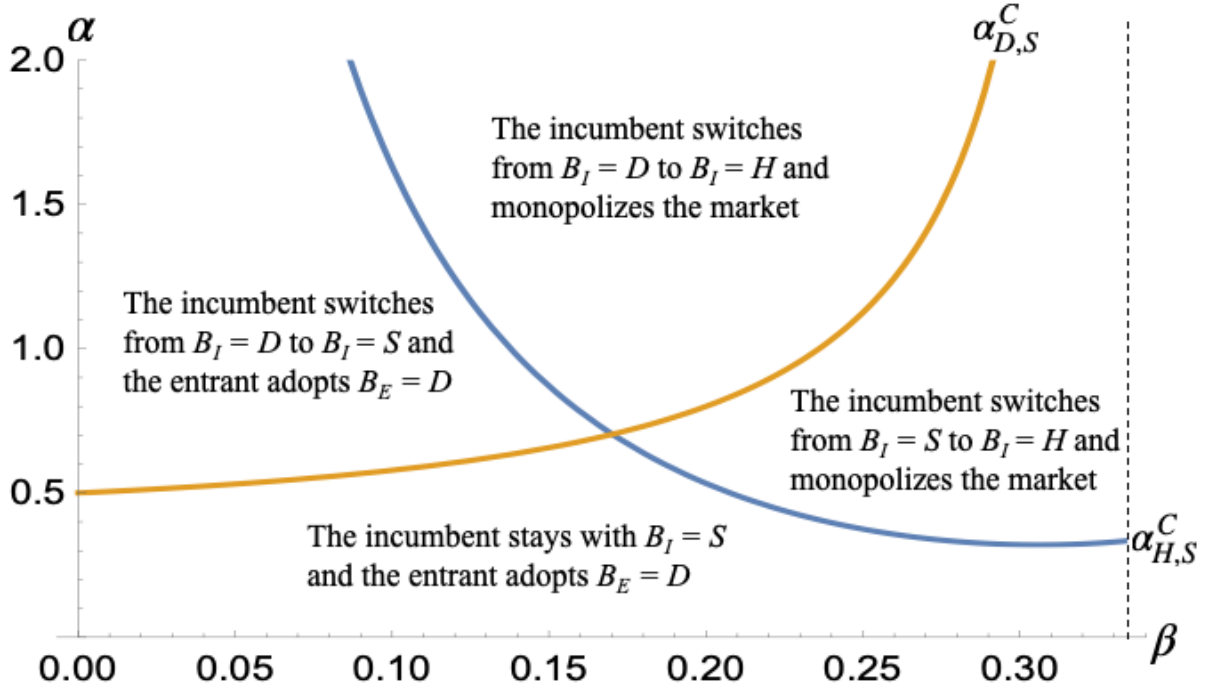
Figure 2: $\alpha_{D,S}^C$ and $\alpha_{H,S}^C$ as a function of β and the equilibrium business models

To see more explicitly how the platform's ability to adopt the hybrid model affects the equilibrium business models, Figure 2 illustrates the threshold value $\alpha_{H,S}^C$ and $\alpha_{D,S}^C$ as a function of β. It is possible to see that there are 4 regions of interest. When α is small such that $\alpha < \min\left\{\alpha_{H,S}^C, \alpha_{D,S}^C\right\}$, the incumbent adopts the subscription-based model for all β, and the platform's ability to adopt the hybrid model does not change this strategy. Here,

17

the data's commercial benefit is low, so it is unprofitable for the incumbent to adopt any of the two business models the relay on commercializing the users' data.

When α is high such that $\alpha > \max\left\{\alpha^C_{H,S}, \alpha^C_{D,S}\right\}$, the incumbent adopts the data-based model if the hybrid model is not possible, but switches to the the hybrid model if possible. Here, the hybrid model is preferable to both the data-based model and the subscription-based model because it enables the incumbent to both commercialize the users data and monopolize the market by allowing data-sensitive users to choose the subscription plan.

When α is intermediate and β is high such that $\alpha^C_{H,S} < \alpha < \alpha^C_{D,S}$, the incumbent switches from the subscription-based model to the hybrid model. Again, the hybrid model is preferable to both the subscription-based model and the data-based model because it enables the incumbent to "benefit from both worlds" and dominate the market.

A less intuitive region is when α is intermediate and β is low such that $\alpha^C_{D,S} < \alpha < \alpha^C_{H,S}$. Here, when platforms do not have the ability to adopt the hybrid model, the incumbent chooses the data-based model. Yet, counterintuitively, when platforms can adopt the hybrid model, the incumbent's strategy changes not to the hybrid model, but rather to the subscription-based model. The intuition for this result is that in this region the incumbent would have preferred to stick to the data-based model. Yet, if the incumbent does so, the entrant would respond by adopting the hybrid model and would monopolize the market. Given that the data-base model is no longer profitable for the incumbent, the incumbent switches to the subscription-based model. That is, here, it is the threat of competition together with the availability of the hybrid model that incentives the incumbent to choose a privacy focused business model.

Next, consider the case where $\frac{1}{3} < \beta < 1 - v$. If the incumbent adopts the data-based model, then the incumbent loses the market if the entrant adopts either the subscription-based or the hybrid models. Hence, it is never optimal for the incumbent to adopt the data-based model. Again, we are left with the options of adopting the hybrid or the subscription-based model. The incumbent prefers the first option if:

$$\pi_I(H, S) = (1 + \alpha - \beta)\beta > \beta = \pi_I(S, D) \iff \alpha > \beta.$$

As in the case of $\beta < \frac{1}{3}$, the incumbent adopts the hybrid model if the data's commercial benefit is high (above some threshold). Recall that if platforms cannot adopt the hybrid model, then when $\beta > \frac{1}{3}$, the incumbent adopts the subscription-based model. Hence, the availability of the hybrid model changes the incumbent's business model from the subscription-based to the hybrid model when α is high.

Comparison with the incumbent's business models when the incumbent is a monopoly

In this subsection we study how competition affects the platforms' business models, when platforms can adopt the hybrid model.

 We start by solving for the incumbent's profit from adopting the hybrid model, when the incumbent is a monopoly. The incumbent's monopolistic maximization problem is similar to the maximization problem under competition (as described in equation (7), with the exception that now the users' alternative utility is 0 instead of v. We therefore have:

Proposition 5. *Suppose that the incumbent is a monopoly that adopts the hybrid model. Then, the incumbent charges and earns*

$$p_I(H) = \begin{cases} \frac{1+\alpha}{2}, & \text{if } \alpha < 2(\beta + v) - 1, \\ \beta + v, & \text{if } \alpha > 2(\beta + v) - 1, \end{cases} \tag{9}$$

$$\pi_I(H) = \begin{cases} \frac{(1+\alpha)^2}{4}, & \text{if } \alpha < 2(\beta + v) - 1, \\ (1 + \alpha - \beta - v)(\beta + v), & \text{if } \alpha > 2(\beta + v) - 1. \end{cases} \tag{10}$$

 The intuition behind this result is as follows. When α is small, p_I is increasing with α because the incumbent takes advantage of the high commercial value of data and sways users to choose the data-plan over the subscription-plan by charging a higher price for the subscription plan. Once α reaches $2(\beta+v)-1$, the utility that users that join the subscription-plan receive reaches 0. In this case, the incumbent extracts all of the utility users that join the subscription-plan enjoy $(\beta + v)$, and the incumbent cannot keep increasing the price (as a function of α).

 Notice that an incumbent that adopts the hybrid model charges a higher price under monopoly than under competition.[9] Next, we turn to compare the incumbent's profit in the three business models, given that the incumbent is a monopoly that can adopt the hybrid model:

Proposition 6. *Suppose that the incumbent is a monopoly that can adopt $B_I = \{S, D, H\}$. Then, the incumbent adopts the hybrid model if $\alpha > \alpha_{H,S}^M = \beta + v$ and adopts the subscription-based model otherwise, where $\alpha_{H,S}^M > 2(\beta + v) - 1$. The incumbent never adopts the data-based model.*

 Intuitively, as under competition, under monopoly the hybrid model is always more profitable than the data-based model. Moreover, the hybrid model is profitable when data has

---

[9]To see why, we have that $\frac{1+\alpha}{2} > \beta$ whenever $0 < \alpha < 2(\beta+v)-1$ and $\beta+v > \beta$ whenever $\alpha > 2(\beta+v)-1$

high commercial value while the subscription-based model is more profitable otherwise. Yet, the two threshold values of α under monopoly and under competition are different. Figure
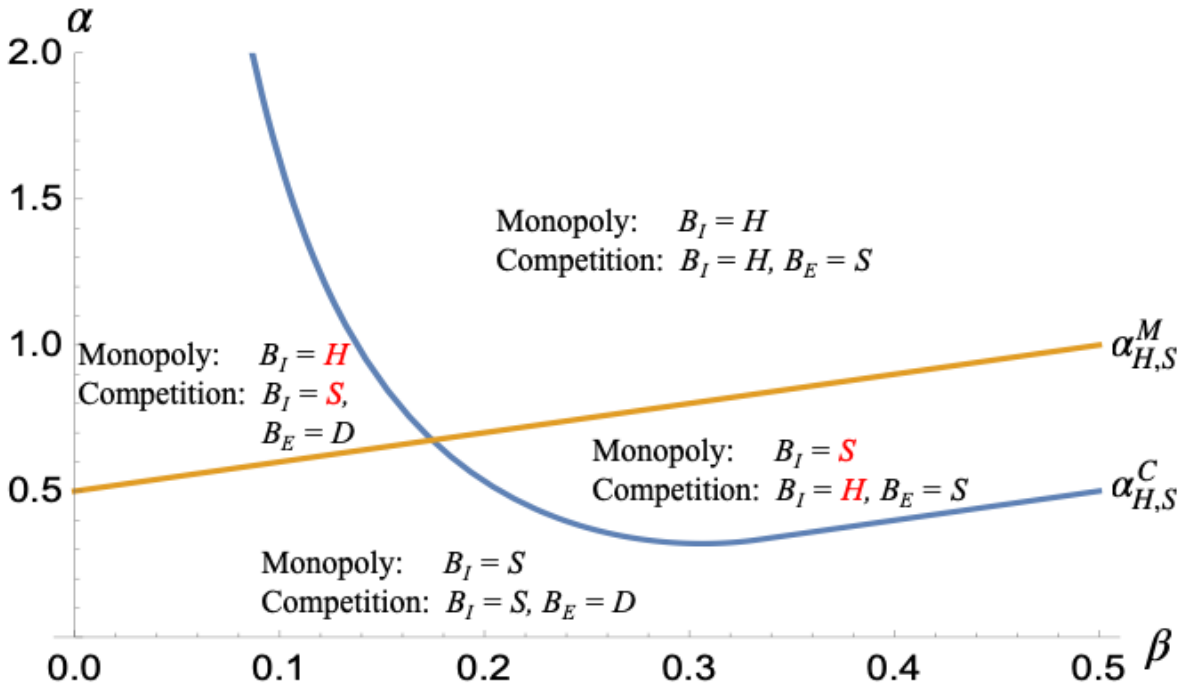
Figure 3: $\alpha_{H,S}^M$ and $\alpha_{H,S}^C$ as a function of β and the equilibrium business models (for $v = \frac{1}{2}$)

3 illustrates the two thresholds, $\alpha_{H,S}^M$ and $\alpha_{H,S}^C$ as a function of β, where recall that $\alpha_{H,S}^C$ is equal to 8) when $\beta < \frac{1}{3}$ and equals to β when $\beta > \frac{1}{3}$. The figure shows that for low values of α, such that $\alpha < \min\left\{\alpha_{H,S}^M, \alpha_{H,S}^C\right\}$, the incumbent adopts the subscription-based model and competition does not change the incumbent's choice. Intuitively, for low commercial value of data, it is optimal to avoid commercializing the users data and instead charge users for the value generated by the platform. For the opposite reason, the incumbent adopts the hybrid model under both monopoly and competition when α is very high, such that $\alpha > \max\left\{\alpha_{H,S}^M, \alpha_{H,S}^C\right\}$. Yet, competition changes the incumbent's business model for inter-mediate values of α, when β is either high or low. For low values of β and intermediate values of α, such that $\alpha_{H,S}^M < \alpha < \alpha_{H,S}^C$, competition motivates the incumbent to switch from the hybrid model to the subscription-based model. Here, competition makes the hybrid model less profitable for the incumbent, who, under the hybrid model, needs to compete with the entrant on all users. As β is small, the incumbent is better off switching to the subscription-based model that focus on data-sensitive users, and share the market with the entrant, who focuses on data-insensitive users. The opposite case occurs for high values of β and intermediate values of α, such that $\alpha_{H,S}^C < \alpha < \alpha_{H,S}^M$. Now, the competition motivates the incumbent to switch from the subscription-based model to the hybrid model. Here, under

monopoly, it is better to take advantage of the high β in order to sell to all users without commercializing their data. Yet, if the incumbent will do so under competition, the entrant can adopts the data-based model and steal the data-insensitive users from the incumbent. Anticipating this, the incumbent adopts the hybrid model and takes advantage of the high β to monopolize the market.

Does the hybrid model implement the welfare maximizing outcome?

In this subsection we comment on whether the hybrid model implements the welfare-maximizing outcome. To maximize welfare, all users must join the same platform, in order to benefit from the data's public benefit. As for data commercialization, it is welfare enhancing to only commercialize data from users with $k < \alpha$. The first two business models that our paper considers, the data-based and subscription-based models, cannot implement the first-best outcome either because the market is not fully covered by the same platform, or because the market is fully covered by the incumbent, but the incumbent commercializes data from all users who join it or from none of the users. The hybrid model enables the incumbent to implement the first-best outcome because the incumbent serves all users and only commercializes data of users with $k < p_I$. Therefore, the incumbent would implement the welfare-maximizing outcome when $p_I = \alpha$, yet would over-commercialize (under-commercialize) data when $p_I > \alpha$ ($p_I < \alpha$). The following proposition shows how over and under commercialization depend on the model's parameters:

Proposition 7. *Suppose that the incumbent adopts the hybrid model. Then, the incumbent commercializes more data under monopoly than under competition. Moreover, in comparison with the welfare-maximizing outcome:*

(i) *When, $\alpha < 2(\beta + v) - 1$, a monopolistic incumbent over-commercializes data and a competitive incumbent over-commercializes (under-commercializes) data when β is high (low).*

(ii) *When $\alpha > 2(\beta + v) - 1$, a monopolistic incumbent over-commercializes (under-commercializes) data when $\beta + v$ is high (low), and a competitive incumbent always under commercializes data.*

Intuitively, a monopolistic incumbent charges a higher price for the subscription-plan than a competitive incumbent, hence derives more users to join the data-plan than a competitive incumbent. When data has low commercial benefit (part (i)), the monopolistic incumbent charges a price that is higher than α, and thus over-commercializes data, relative to the

21

welfare maximizing outcome. It is easy to understand the intuition, when looking at the case where $\alpha \to 0$. In this case $p_I \to 1/2$, which is the standard monopolistic price. Given the low α, we get that $p_I > \alpha$. In contrast, a competitive incumbent who sets a lower price may over or under commercialize data. Following this same logic, when data has high commercial benefit (part (ii)), a competitive incumbent charges a low price such that it under-commercializes data. In contract, a monopolistic incumbent who charges a higher price may over or under commercialize data.

6   Managerial Implications

In today's information age, where data plays an increasingly important role in platforms' value creation, platforms are faced with the value capture dilemma of whether to base their business model on the "traditional" practice of charging users for their services, adopt the new business model of commercializing the users' data, or do both. Our analysis provides guidelines with respect to when it is optimal for platforms to adopt each business model, and thus has important managerial implications both for monopolistic and competing platforms.

First and foremost, we find that platforms' optimal business model should consider not only the commercial value of data but also the public benefit of data. While it is reasonable to expect that commercializing data is the profitable business model if the commercial benefit of data is high, our model reveals that for intermediate commercial value, the magnitude of the data's public benefit is crucial for determining the optimal business model. That is, it is imperative for platforms to gauge what benefit the data they collect on users provides to other users. For example, in the case of a navigation app such as Waze which collects information on drivers' location, the data collected is crucial to other users that use the app and, in fact, is the core of the service that the app provides. Hence, in this case, the public benefit of data is expected to be high. In contrast, the public benefit of data of apps like Ride with GPS that provide route directions to cyclists is relatively low. While the app also collects data on the rider's location, the data is mainly used to provide the rider with directions (rather than provide congestion information) and thus does not require up-to-date data on the location of other riders, as in the case of Waze.

Our analysis finds that, if a hybrid model is not available (e.g., banned by regulation, too complex to implement, or not popular as a business model), then in the case of a monopolistic platform,) and intermediate commercial value, platforms should go with the data-based model when the public benefit of data is either small or large. For intermediate values of the public benefit of data, it is the subscription-based model that maximizes profits. For example, when Facebook and Netflix first launched, the commercial value on both platforms

was of intermediate value. For Netflix the public benefit of data was likely medium as the data was mostly used to help recommending users shows that would be of good fit with their preferences, yet in the case of Facebook it is the actual presence and information of other users that created value (and thus the public benefit of data was likely high). Indeed, Facebook went with a data-based model while Netflix with the subscription option.

Under the threat of competition, an incumbent should change its business model under the threat of entry, only if the commercial benefit of data is of intermediate value. In this case, if the public benefit of data is high, an incumbent platform should switch from the data-based model to the subscription-based one. The data-based model makes the platform vulnerable to entry by a platform that does not commercialize users' data. In contrast, the subscription-based model enables the platform to gain a higher market share and if the public benefit of data is high, to dominate the market. Platforms should follow the opposite strategy (switch from subscription- to data-based model), if the public benefit of data is low. The entrant platform, then, should enter with a business model that is the opposite to the one offered by the incumbent. For example, when Hulu entered the market to compete with Netflix (which at the time offered only subscription plans), it entered with a data-based business model.

When the hybrid model is possible, the hybrid model always dominates the data-based model, because it allows a platform to attract all users in the market where the data-based model results in partial market coverage. For example, Meta recently launched in Europe the hybrid model that includes a free service with data-tracking and advertising and a no-ads privacy subscription service for a fee. Pandora entered the music streaming market with a hybrid business model that allowed users to choose between an ad-based model and a subscription one. The one exception is the case where the commercial value of data is low, in which case a subscription business model is more profitable than the hybrid model.

More importantly, under competition the hybrid business model can help incumbent platforms to deter entry, or at least prevent the entrant from dominating the market. Specifically, when the public benefit of data is high, incumbent platforms should choose the hybrid model as it would deter entry. When the public benefit of data is low and the commercial value of data is high, the incumbent platform should choose the subscription-based model, despite the higher profitability of the data-based model. Going with a data-based model would result in the entrant entering with a hybrid model and monopolizing the market. The incumbent, thus, is better off choosing the subscription model and sharing the market with the entrant.

7 Conclusion

Data is becoming an essential asset for platforms and an important determinant of platforms' monetization strategies. We develop a tractable model to study how competition affects platforms' optimal business model in a market where data collected on users helps platforms improve the quality of their service. Platforms can choose between three business models: data-based, subscription-based, and hybrid. We find that the effect of competition on platforms' optimal business model depends on the interaction between the public and commercial benefit of data. Importantly, we show that when the public benefit of data is low, competition may in fact drive platforms to choose to commercialize users' data rather than keep it private. That is, competition does not necessarily promote a more privacy-sensitive market.

Our second main result concerns with the hybrid business model. Allowing platforms to discriminate across users based on whether they share their data for commercialization or not–i.e., to offer the hybrid model–may incentivize an incumbent platform to choose a more privacy-sensitive business model. Still, allowing for the hybrid model may result in a more concentrated market where the incumbent can deter the entry of a new platform into the market.

References

[1] Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. 2022. "Too Much Data: Prices and Inefficiencies in Data Markets." *American Economic Journal: Microeconomics*, 14 (4): 218-56.

[2] Aridor, Guy, Yeon-Koo Che, and Tobias Salz. 2023. "The effect of privacy regulation on the data industry: Empirical evidence from GDPR." *The Rand Journal of Economics* 54(4): 695-730.

[3] Bergemann, Dirk, Alessandro Bonatti and Tan Gan. 2022. "The Economics of Social Data." *The Rand Journal of Economics* 53(2): 263-296.

[4] Biglaiser, Gary, and Jacques Crémer. 2020. "The value of incumbency when platforms face heterogeneous customers." *American Economic Journal: Micro* 12(4):1-43.

[5] Caillaud, Bernard, and Bruno Jullien. 2001. "Competing cybermediaries." *European Economic Review* 45 (4-6): 797–808.

[6] Caillaud, Bernard, and Bruno Jullien. 2003. "Chicken & egg: Competition among intermediation service providers." *The RAND Journal of Economics* 34 (2): 309–328.

[7] Casadesus-Masanell, Ramon, and Feng Zhu. 2010. "Strategies to Fight Ad-sponsored Rivals." Management Science 56(9): 1484–1499.

[8] Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim. 2019. "Privacy and personal data collection with information externalities." *Journal of Public Economics* 173:113-124.

[9] Fairfield, Joshua AT, and Christoph Engel. 2015. "Privacy as a public good." Duke Law Journal, 65: 385.

[10] Fainmesser, Itay, Andrea Galeotti, and Ruslan Momot. 2022. "Digital Privacy." Management Science 69(6): 3157-3173.

[11] Hagiu, Andrei. 2006. "Pricing and commitment by two-sided platforms." *The RAND Journal of Economics* 37 (3): 720–737.

[12] Hagiu Andrei and Julian Wright. 2023. "Data-enabled learning, network effects and competitive advantage." *RAND Journal of Economics* 54(4): 638-667.

[13] Hałaburda, Hanna, and Yaron Yehezkel. 2013. "Platform competition under asymmetric information." *American Economic Journal: Microeconomics* 5 (3): 22–68.

[14] Hałaburda, Hanna, and Yaron Yehezkel. 2016. "The role of coordination bias in platform competition." *Journal of Economics and Management Strategy* 25 (2): 274–312.

[15] Hałaburda, Hanna, and Yaron Yehezkel. 2019. "How beliefs affect platform competition." *Journal of Economics and Management Strategy* 28 (1), 49-49.

[16] Hałaburda, Hanna, Bruno Jullien, and Yaron Yehezkel. 2020. "Dynamic platform competition: how history matters?" *The RAND Journal of Economics* 51 (1): 3-31.

[17] Johnson Garrett A., Scott K. Shriver, and Samuel G. Goldberg. 2023. "Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR." *Management Science.*

[18] Janßen, Rebecca, Reinhold Kesler, Michael E. Kummer, and Joel Waldfogel. 2022. "GDPR and the Lost Generation of Innovative Apps." NBER Working Paper 30028.

[19] Jullien, Bruno. 2011. "Competition in multi-sided markets: Divide and conquer." *American Economic Journal: Microeconomics* 3 (4): 186–220.

[20] Jullien, Bruno, Alessandro Pavan, and Marc Rysman. 2021. "Two-sided markets, pricing, and network effects." In Handbook of Industrial Organization, vol. 4, no. 1, pp. 485-592. Elsevier.

[21] Katz, Michael L., and Carl Shapiro. 1986. "Technology adoption in the presence of network externalities." *Journal of Political Economy* 94 (4): 822-841.

[22] Ke, Tony T., and K. Sudhir. 2022. "Privacy rights and data security: GDPR and personal data driven markets. *Management Science.*

[23] Liang, Annie, and Erik Madsen. 2019. "Data Sharing and Incentives." Available at SSRN 3485776.

[24] Markovich, Sarit, and Yaron Yehezkel. 2022. "Group Hug: Platform Competition with User-groups." *American Economic Journal: Micro,* 14 (2): 139-175.

[25] Markovich, Sarit, and Yaron Yehezkel. 2023, "For the public benefit": data policy in platform markets". Working paper.

[26] O'Brien, Daniel and Doug Smith. 2014. "Privacy in online markets: A welfare analysis of demand rotations." Working Paper No. 323

Appendix A

Below are the proofs for all lemmas and propositions in the text.

Proof of Proposition 1:

Evaluated at $\beta = 0$, The gap $\alpha_{D,S}^M - \alpha_{D,S}^C = \frac{1}{2} > 0$. For $\beta \to \frac{1}{3}$, $\alpha_{D,S}^C \to \infty$ while $\alpha_{D,S}^M$ is finite, hence $\alpha_{D,S}^M - \alpha_{D,S}^C < 0$. Moreover, the gap $\alpha_{D,S}^M - \alpha_{D,S}^C$ is strictly concave β because:

$$\frac{\partial \left(\alpha_{D,S}^M - \alpha_{D,S}^C\right)}{\partial \beta} = -\frac{4}{(1-3\beta)^3} - \frac{2}{v} < 0,$$

implying that there is a unique solution to $\alpha_{D,S}^M = \alpha_{D,S}^C$ at:

$$\overline{\beta} = \frac{1}{12}(2 - 5v + \sqrt{4 + v(4 + 25v)}).$$

Finally, we have that $\overline{\beta} < 1 - v$ if $v < 3 - \sqrt{5} \cong 0.764$.

Proof of Proposition 2:

The gap $n_I(D) - n_I(D, S)$ satisfies

$$n_I(D) - n_I(D, S) = \frac{v}{1-\beta} - \left(1 - \frac{1-\beta}{2(1-2\beta)}\right) >$$

$$\frac{\frac{1}{2}}{1-\beta} - \left(1 - \frac{1-\beta}{2(1-2\beta)}\right) = \frac{(2-3\beta)\beta}{2(1-3\beta+2\beta^2)} > 0,$$

where the first inequality follows because $v > \frac{1}{2}$ and the second inequality follows because $\beta < \frac{1}{3}$.

Next, consider the gap $p_I(S) - p_I(S, D)$. We have:

$$p_I(S) - p_I(S, D) = v + \beta - \frac{1-\beta}{2} > \frac{1}{2} + \beta - \frac{1-\beta}{2} = \frac{3}{2}\beta > 0,$$

where the first inequality follows because $v > \frac{1}{2}$.

Proof of Proposition 3:

Suppose that the incumbent chooses $B_I = D$ and the entrant chooses $B_E = H$. The entrant's problem is to set p_E to maximize:

$$\max_{p_E} \pi_E(p_E|(D, H)) = \alpha p_E + (1 - p_E)p_E, \qquad (11)$$

$$\text{s.t.} \quad \beta + v - p_E \geq \max\{v - p_E, 0\} \text{ and } p_E \leq 1.$$

The first constraint ensures that the user with $k = p_E$ who is indifferent between the entrant's data plan and the subscription plan prefers to join the entrant over joining the incumbent's data plan or stay out of both platforms. The second constraint ensures that the indifferent users with $k = p_E$ has an internal solution.

The solution to the unconstraint problem is $p_E = \frac{1+\alpha}{2}$, which satisfies the constraint $p_E < 1$ if $\alpha < 1$. Moreover, p_E always satisfy the constraint $\beta + v - p_E \geq v - p_E$ and the constraint $\beta + v - p_E \geq 0$ requires that $\alpha < 2(\beta + v) - 1$, where $0 < 2(\beta + v) - 1 < 1$ because $v > 1/2$ and $\beta < 1 - v$. Hence, for $0 < \alpha < 2(\beta + v) - 1$, the entrant sets $p_E = \frac{1+\alpha}{2}$ and earns $\pi_E(D, H) = \frac{(1+\alpha)^2}{4}$. For $2(\beta + v) - 1 < \alpha$, there is a corner solution with $b + v - p_E = 0$, or $p_E = \beta + v < 1$. The entrant sets in this case $p_E = \beta + v$ and earns $\pi_E(D, H) = (\beta + v)(1 + \alpha - \beta - v)$. In both cases, the incumbent earns $\pi_I(D, H) = 0$.

Next, suppose that the entrant chooses $B_E = S$. From the analysis of Section 4, when $\beta < 1/3$, the entrant earns in this case $\pi_E(D, S) = \frac{(1-\beta)^2}{4(1-2\beta)}$. When $0 < \alpha < 2(\beta + v) - 1$, the entrant prefers the hybrid model if $\pi_E(D, H) > \pi_E(D, S)$, or $\alpha > \bar{\alpha} = \frac{1-\beta}{\sqrt{1-2\beta}} - 1$, where

$$2(\beta + v) - 1 - \left(\frac{1-\beta}{\sqrt{1-2\beta}} - 1\right) > 2(\beta + \frac{1}{2}) - 1 - \left(\frac{1-\beta}{\sqrt{1-2\beta}} - 1\right) = 1 + 2\beta - \frac{1-\beta}{\sqrt{1-2\beta}} > 0,$$

where the first inequality follows because $v > \frac{1}{2}$ and the second inequality follows when $\beta < 1/3$. Again from the analysis of Section 4, when $\beta > 1/3$, the entrant earns in this case $\pi_E(D, S) = \beta$. When $0 < \alpha < 2(\beta + v) - 1$, the entrant prefers the hybrid model if $\pi_E(D, H) > \pi_E(D, S)$, or $\alpha > \bar{\alpha} = 2\sqrt{\beta} - 1$, where it is possible to show that $2\sqrt{\beta} - 1 < 2(\beta + v) - 1$ and $2\sqrt{\beta} - 1 = \frac{1-\beta}{\sqrt{1-2\beta}} - 1$ at $\beta = 1/3$.

A third option for the entrant is to adopt the data-based model. Yet, when both platform adopt the same business model, the incumbent wins the market due to its focal position while the entrant earns 0.

To summarize, we have that when $0 < \alpha < \bar{\alpha}$, where

$$\bar{\alpha} = \begin{cases} \frac{1-\beta}{\sqrt{1-2\beta}} - 1, & \text{if } \beta < \frac{1}{3}, \\ 2\sqrt{\beta} - 1, & \text{if } \beta > \frac{1}{3}, \end{cases}$$

and $\bar{\alpha} < 2(\beta + v) - 1$, the entrant responds by adopting the subscription-based model. The two platforms earn $\pi_I(D, S)$ and $\pi_E(D, S)$ as defined in Section 4. When $\bar{\alpha} < \alpha < 2(\beta+v)-1$, the entrant adopts the hybrid model, charges $p_E = \frac{1+\alpha}{2}$ and earns $\pi_E(D, H) = \frac{(1+\alpha)^2}{4}$. When $2(\beta+v)-1 < \alpha$, the entrant adopts the hybrid model and there is a corner solution in which the entrant sets $p_E = \beta + v$ and earns $\pi_E(D, H) = (\beta + v)(1 + \alpha - \beta - v)$. In both cases, the

28

incumbent earns $\pi_I(D, H) = 0$.

Proof of Proposition 4:

Suppose that $\beta < \frac{1}{3}$. We first compare between the incumbent's profit when it adopts $B_I = D$ and the entrant responds by adopting $B_E = S$ (which occurs only when $\alpha < \overline{\alpha}$). We have that $\pi_I(H, S) > \pi_I(D, S)$ if:

$$\pi_I(H, S) = (1 + \alpha - \beta)\beta > \frac{\alpha(1 - 3\beta)}{2(1 - 2\beta)} = \pi_I(D, S), \tag{12}$$

$$\Downarrow$$

$$\alpha < \alpha_{D,H}^C = \frac{2\beta(1 - 2\beta)}{1 - 4\beta}.$$

Yet, $\alpha_{D,H}^C > \overline{\alpha}$, implying that whenever adopting $B_I = D$ motivates the entrant to adopt $B_E = S$ (which occurs when $\alpha < \overline{\alpha}$), it is not optimal for the incumbent to adopt $B_I = D$, as the incumbent prefers $B_I = H$ over $B_I = D$. We are therefore left with two options, either setting $B_I = H$ or setting $B_I = S$. The incumbent prefers the first option when:

$$\pi_I(H, S) = (1 + \alpha - \beta)\beta > \frac{(1 - \beta)^2}{4(1 - 2\beta)} = \pi_I(S, D), \tag{13}$$

$$\Downarrow$$

$$\alpha > \alpha_{H,S}^C = \frac{(1 - \beta)(1 - 5\beta + 8\beta^2)}{4\beta(1 - 2\beta)},$$

where $\alpha_{H,S}^C$ is decreasing with β.

Proof of Proposition 5:

The incumbents' problem in the hybrid model when the incumbent is a monopoly is:

$$\max_{p_I} \pi_I(p_I|(H)) = \alpha p_I + (1 - p_I)p_I, \tag{14}$$

$$\text{s.t.} \quad \beta + v - p_I \geq 0 \text{ and } p_I \leq 1.$$

The unconstrained solution is $p_I = \frac{1 + \alpha}{2}$. Notice first that $p_I < 1$ if $\alpha < 1$. Moreover, at this price, users gain non-negative utility if:

$$\beta + v - \frac{1 + \alpha}{2} \geq 0 \iff \alpha < 2(\beta + v) - 1,$$

where $2(\beta + v) - 1 < 1$ because $\beta < 1 - v$. Hence, we have that for $\alpha < 2(\beta + v) - 1$,

$p_I = \frac{1+\alpha}{2} < 1$ and the incumbent earns $\pi_I(H) = \frac{(1+\alpha)^2}{4}$. Next, suppose that $\alpha > 2(\beta+v)-1$. In this case, the constraint $\beta+v-p_I \geq 0$ binds. Therefore, $p_I = \beta+v < 1$ and the incumbent earns $\pi_I(H) = (1+\alpha-\beta-v)(\beta+v)$.

Proof of Proposition 6:

We first show that the incumbent always prefers the hybrid model over the data-based model. When $\alpha < 2(\beta+v)-1$, we have:

$$\pi_I(H) - \pi_I(D) = \frac{(1+\alpha)^2}{4} - \frac{v\alpha}{1-\beta} > \frac{(1+\alpha)^2}{4} - \alpha = \frac{(1-\alpha)^2}{4} > 0,$$

where the first inequality follows because $v < 1-\beta$. When $\alpha > 2(\beta+v)-1$, we have:

$$\pi_I(H) - \pi_I(D) = (1+\alpha-\beta-v)(\beta+v) - \frac{v\alpha}{1-\beta} = \frac{(1-\beta-v)(v+\beta(1+\alpha-\beta-v))}{1-\beta}$$

$$> \frac{(1-\beta-v)(v+\beta(\beta+v))}{1-\beta} > 0,$$

where the first inequality follows because $\alpha > 2(\beta+v)-1$ and the second inequality follows because $v < 1-\beta$.

We are therefore left with the comparison between $\pi_I(H)$ and $\pi_I(S)$. When $\alpha < 2(\beta+v)-1$, we have:

$$\pi_I(H) - \pi_I(S) = \frac{(1+\alpha)^2}{4} - (\beta+v) < -(1-\beta-v)(\beta+v) < 0,$$

where the first inequality follows because $\alpha < 2(\beta+v)-1$ and the second inequality follows because $v < 1-\beta$. Hence, it is optimal to adopt the subscription-based model when $\alpha < 2(\beta+v)-1$. When $\alpha > 2(\beta+v)-1$, we have:

$$\pi_I(H) - \pi_I(S) = (1+\alpha-\beta-v)(\beta+v) - (\beta+v) > 0 \iff \alpha > \beta+v,$$

where $\beta+v > 2(\beta+v)-1$ because $0 < \beta < 1-v$. Hence, there is a threshold, $\alpha_{H,S}^M = \beta+v$, where $\alpha_{H,S}^M > 2(\beta+v)-1$, such that the monopolistic incumbent adopts the subscription-based model if $\alpha < \alpha_{H,S}^M$ and the hybrid model otherwise.

Proof of Proposition 7:

Suppose that the incumbent adopts the hybrid model. Recall that p_I under competition is lower than under monopoly, hence, there is less data commercialization under competition

than under monopoly.

Next, consider part (i). When $0 < \alpha < 2(\beta + v) - 1$, the monopoly sets $p_I = \frac{1+\alpha}{2}$. We have that $\frac{1+\alpha}{2} > \alpha$, hence, there is over-commercialization of data under monopoly because:

$$\frac{1+\alpha}{2} - \alpha = \frac{1-\alpha}{2} > \frac{1 - (2(\beta + v) - 1)}{2} = 1 - \beta - v > 0,$$

where the first inequality because $\alpha < 2(\beta + v) - 1$ and the second inequality follows because $\beta < 1 - v$. A competitive incumbent sets $p_I = \beta$, which can be lower or higher than α. This is because $\beta < 2(\beta + v) - 1$ (which follows because $2(\beta + v) - 1 - \beta = 2v - 1 + \beta > 0$, where the inequality follows because $v > 1/2$).

Next, consider part (ii). When $2(\beta + v) - 1 < \alpha$, the monopoly sets $p_I = \beta + v$. We have that $\beta + v$ can be higher or lower than α, because $\beta + v > 2(\beta + v) - 1$ (which follows because $\beta + v - (2(\beta + v) - 1) = 1 - \beta - v > 0$. A competitive incumbent sets $p_I = \beta$, which is lower than α, hence, there is under-commercialization, because: $\alpha - \beta > 2(\beta + v) - 1 - \beta = 2v - 1 + \beta > 0$, where the first inequality follows because $\alpha > 2(\beta + v) - 1$ and the second inequality follows because $v > 1/2$.