Take Out Your Data: Impact of the Right to Data Portability on Platform Competition*

Hooman Hidaji

University of Calgary, hooman.hidaji@haskayne.ucalgary.ca

The interactive Internet, smartphones, and the Internet of Things have made customers producers of data as well as users of it. Firms and platforms have leveraged this customer-generated data to provide services that are customized to each customers' needs. It is difficult for new entrants to provide equivalent services due to the lack of insight from customer data. Regulators are wary of the implications of this data-driven barrier for competition. One suggested solution, which is enforced in the recent European Union's General Data Protection Regulation (GDPR), obliges firms to allow customers to download their data and potentially transfer or port it to other firms or platforms. This is known as the right to data portability (RDP). It has been suggested that data portability is beneficial to competition. However, there is no empirical or analytical evidence to support this claim. In this study, we use a two-stage model to analyze the interaction of customers, platforms, and regulators with respect to RDP, and their impact on pricing, competition, and customers. Our findings suggest that enforcing the right to data portability may reduce the demand and profitability of new firms in some situations by increasing price and reducing the demand and profit for new entrants. Specifically, if customers' valuation for the platforms is moderate compared to the effectiveness of platforms in leveraging the customer data and data porting is incompatible, enforcing RDP may drive the incumbent to reduce its price and increase its market share at the cost of the entrant's market share. Moreover, enabling data portability may reduce customer surplus due to increased price of entrant platform and reduced demand for the incumbent. The important implication of this for regulators is that enforcing RDP is not always beneficial to the competition or customers.

Keywords: right to data portability, platform competition, customer-generated data, data privacy regulation

---

* Work-in-progress, please do not cite or quote without the author's permission. Feedback is welcome.

1. Introduction

The interactive Internet, smartphones, and the Internet of Things (IoT) have made customers producers of data as well as users of it. Firms and platforms have leveraged this customer-generated data to provide services that are customized to each customers' needs and are, therefore, appealing to them. For example, Google Search improves as it learns about a customers' characteristics such as their location and interests; Netflix uses customer behavior data to recommend content to customers[1]; using a web browser becomes easier once the form information and passwords are stored in the browser; and mobile platforms provide customized services based on the data generated when customers use these platforms. It is difficult for new entrants in these industries to provide equivalent services due to the lack of insights from customer data, and therefore, customers are disincentivized from switching to a new entrant. Thus, the lack of access to customer data acts as a barrier to entry for new entrants. Regulators are wary of the implications of this data-driven barrier for competition. One suggested solution obliges firms to allow customers to download their data, and potentially transfer or port it to other firms or platforms. This is known as the right to data portability (RDP), and is the topic of this study.

The recent General Data Protection Regulation (GDPR) from the European Union (EU), which went into effect in May 2018, enforces RDP. This has compelled international companies such as Google, Apple, and Facebook to allow customers to download their data (Conger 2018). Customers can then port their data to another platform. Data portability is not limited to personal data for individuals, and can be extended to business applications and enterprise systems such as Google's G Suite and Microsoft's Office 365. While the technical details of porting between platforms is yet to be established (some initiatives include Data Transfer Project[2] and Universal Digital Profile[3]), the European Commission's European Political Strategy Centre suggests that data portability

[1] which reportedly drives more than 80 percent of customer activity on the platform: https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like

[2] https://datatransferproject.dev

[3] https://techcrunch.com/2018/05/22/the-birth-of-the-universal-digital-profile/

is beneficial to competition (EPSC 2017, p.11). This assertion is the basis of RDP regulation. Crucially, however, there is no empirical or analytical evidence to support this claim. It is, therefore, critical to study the implications of RDP for pricing and competition, as well as for customer-generated data, and customer surplus. We set out to study these implications in this study. Due to the broad applicability of data privacy regulation such as the GDPR, and the range of companies collecting customer data, the understanding of this has vast implications for customers, firms, and regulators in a wide array of industries. Moreover, the GDPR has influenced third-countries (non-EU member states) to update their privacy laws, which extends the reach of such regulations. For example, in Canada, the consultation process on digital and data transformation has started and is expected to result in incorporation of concepts such as RDP in the data privacy acts and regulations[4].

The existence of multiple stake-holders and strategic players, makes the RDP a complex problem. In this paper we consider the impact of data portability on pricing and competition, as well as on customer-generated data and customer surplus. We use a two-stage game among two platforms, competing for customers with heterogeneous preferences for the platforms. We assume an incumbent platform in Stage 1, and a second platform that enters the market in Stage 2. The incumbent can leverage customer-generated data from Stage 1 to create better services for its customers. If RDP is enforced, the entrant platform can too benefit from customer-generated data. Our findings suggest that data portability does not always create more competition, and may even reduce the demand and profit for the new platform in certain situations. Moreover, enabling RDP may reduce customer surplus due to increased price for the entrant platform and reduced demand for the incumbent.

The rest of this study is organized as follows. In the next subsection we present the relevant prior literature. Section 2 provides our model. Section 3 and 4 analyze the two stages of the problem. We present our results in Section 5, and conclude in Section 6.

[4] `https://www.jdsupra.com/legalnews/canada-to-update-data-law-to-gdpr-16052`

1.1. Literature

Despite the importance of the issue, RDP has not received much attention from researchers. There are studies on the technical details of data portability in different platforms, and especially, in social networks (Bojars et al. 2008, Razmerita et al. 2009, Shirazi et al. 2012). On the other hand, there are studies on legal implications of RDP. Swire and Lagos (2012), using arguments in law and regulation, suggest that the right to data portability may reduce consumer surplus. Engels (2016) illustrates the complexity of RDP and suggests that it is very nuanced and should be dealt with in that manner. De Hert et al. (2018) provide a systematic and technical interpretation of the data portability rules within GDPR. From a behavioral point of view, Fialová (2014) suggests that data portability results in informational self-determination, which is associated with information privacy. We do not consider such interactions, instead, focus on the impact of data portability as a data privacy mechanism on competition and other factors, as well as its effect on the customer-generated data.

There are only a few studies to consider the economic impacts of data portability. Christensen et al. (2013) study the impact of data protection regulation, including data portability, on firms within the EU. They consider the costs of compliance and lack of access to data, and estimate a significant negative impact on firms and economy as a result of enforcing these regulations. We do not consider this in our study. In the closest study to ours, Lam and Liu (2018) consider the impact of data analytics and inferred data on how entry is facilitated or inhibited by data portability. They find that when inferred data cannot be ported, as is the case with the current GDPR regulation, data portability facilitates switching. However, if inferred data can be ported in addition to customer data, data portability may hinder customer switching due to the demand-expansion effect, where strategic customers provide more data to the incumbent and strengthen its position and contribute to entry barriers. In our model, we do not consider inferred data, and instead focus on the impact of data portability on switching, competition, and profits of the price-setter incumbent and entrant platforms. Contrary to Lam and Liu (2018), we find conditions under which data portability may inhibit competition by reducing the entrant's demand and profit.

This study contributes to the information privacy literature. Bélanger and Crossler (2011) and Smith et al. (2011) provide surveys of the research in this area. Privacy and its interaction with price discrimination, customer surplus, and competition have been extensively studied (Taylor 2004, Acquisti and Varian 2005, Casadesus-Masanell and Hervas-Drane 2015, Koh et al. 2017). Gal-Or et al. (2018) and Kim (2017) study the impact of privacy on competition due to privacy concerns from targeted advertising. Our paper is different from this stream, in that we consider data portability as a mechanism for customers to take control over their data, and this decision has indirect implications for privacy. Our paper is also related to research on how customers share information with firms. Casadesus-Masanell and Hervas-Drane (2015) consider utility-maximizing customers that disclose information in a retail setting. We use a similar approach in conceptualizing data disclosure by customers, however, our study is different in that we focus on the impact of data portability on stakeholders and outcomes. In the literature, RDP or similar tools have not been considered as mechanisms for customers to manage their privacy and take control of their data.

This paper is also related to the vast literatures on market entry, switching costs, standardization, and product customization. There is an established stream of literature on incumbent and market entry. Our work builds on the seminal work by Salop (1979) where the author considers a circular location for firms and an outside option. Nault and Vandenbosch (2000) consider the timing of entry from entrant with disruptive technologies, and discuss how the incumbent and entrant may act to increase barriers to entry. In the standardization literature, Farrell and Saloner consider standardization as a means to reduce switching costs, but note how it can trap industries in an inferior standard or technology (Farrell and Saloner 1985) and where an installed base may inhibit beneficial innovation (Farrell and Saloner 1986). Dewan et al. (2003) consider the impact of product customization driven from customer data. In contrast, we study the RDP as a data privacy mechanism which has implications for and contributes to literatures on market entry, switching of customers, standardization, and customization.

Caillaud and De Nijs (2014) study a duopoly where firms can price discriminate between old and new customers. We do not consider this type of customer retention strategy, but focus on the customer porting between two platforms. Gehrig and Stenbacka (2007) consider information sharing between two banking systems and its effects on customer poaching in presence of switching costs. We do not consider information sharing by the firms, but where customers can decide to port their data from one firm to another.

Zhu and Iansiti (2012) consider market entry in platform-based markets with customers and developers, considering platform quality, network effects, and consumer expectations. We do not consider two-sided networks in the current study, and extend that paper by considering portability among the two platforms as a form of reducing switching cost and barriers to entry.

Gehrig et al. (2011) consider the ability of firms to use customer data to price-discriminate in a duopoly with incumbent and entrant, where the incumbent is protected due to switching costs. They use a similar vertical differentiation to our study.

Our paper is different from the literature in that when customers are able to port their data to another platform, some of them with strong favor for the entrant platform, have a negative switching cost. While data portability reduces switching cost, it is different from the literature in this field, in that it does not reduce switching cost homogeneously among all customers, but only among customers that are currently using another platform.

We consider market entry in the second stage. Contrary to the prior literature, we consider competition when customers can make a strategic decision on whether to port their data to another platform, and how the decision to enable data portability impacts platform competition and customers.

2. Set-up, Notation and Assumptions

Our model set-up is as follows. We consider a two-stage game between two platforms: the incumbent platform (Incumbent) and the new platform (Entrant). In the first stage, only Incumbent is available. In Stage 2, Entrant enters the market. Each platform sets its own price so as to maximize its profit. All costs are normalized to zero.

Assumption 1 **(Myopic Platform and Customers)**. *Incumbent maximizes its profit in each stage. Customers maximize their utility in each stage.*

Platforms often maximize their current profits. While there is always a chance for a new platform to enter, platforms may ignore predicting entrance in their pricing decisions. In scenarios where demand externalities are important, platforms may reduce their prices initially to gain more demand. However, in this study, we do not consider demand-side externalities. We consider customers to be myopic as well, in that they maximize their utility in each stage.

ASSUMPTION 2 **(Demand)**. *(a) Customers cannot multi-home, i.e. can use at most one platform.*

(b) Customers are homogeneous in terms of their intrinsic valuation for platform use.

(c) Customers have heterogeneous preferences for the two platforms, and are uniformly distributed over $[0, 1]$ *as per the Hotelling model (Hotelling 1929). Incumbent and Entrant are horizontally differentiated and located on 0 and 1 on this line, respectively. Customers incur mismatch cost when they use each of the platforms.*

(d) Customer intrinsic valuation is high enough that the two platforms cover the market in Stage 2.

This implies that the basic functionality of the two platforms is similar. We denote the intrinsic valuation as $v \in R^+$. The customers, however, incur mismatch cost of $t \in R^+$ per unit of distance by using each platform, according to the Hotelling model. Market is not necessarily covered in Stage 1 where there is only one platform. In this model, Entrant always enters the market in Stage 2 and we do not consider market entry; instead, our analysis focuses on demands and profits.

Customers disclose information and generate data by using the services offered by the platforms. We do not differentiate between usage of the platform and data generation. The amount of data generated is denoted as $D \in R^+$. Customers decide how much to disclose based on the trade-off between additional value that they receive from customized service, and the privacy concerns from disclosing the data. We abstract this trade-off as a concave specific form function. The next assumption formalizes this.

Assumption 3 **(Customer Data Disclosure)**. *Customer utility from using the platforms and disclosing data is concave. The amount of data disclosure is decided by the customers to maximize their utility. Customers are homogeneous in their tendency to disclose.*

How much the customers use the platforms and disclose data, is chosen so as to maximize their utility. We assume the form $D[w-D]$ for the utility from data disclosure, where $w \in R^+$ is the tendency of customers to use the platforms and disclose data. Customers are homogeneous in their tendency to disclose. The disclosure can be seen as customers' interaction with the platform, or intensity of use. Using this functional form, the optimal data disclosure for each customer is derived as $D^* = w/2$. This implies that customers' utility improves from disclosing personal data (and the process of using the service) up to a certain threshold, before which the additional value generated is larger than the privacy concerns. Data disclosed beyond this threshold, however, reduces the net utility of customers, due to privacy concerns being higher than additional value generated. Casadesus-Masanell and Hervas-Drane (2015) use a similar form for their data quality construct.

Assumption 4 **(Customer-Generated Data)**. *In Stage 2, platforms may benefit from customer-generated data in Stage 1 as follows:*

(a) Customers of Incumbent in Stage 1 enjoy additional utility through inflated tendency to disclose. This is due to the customized services based on prior data, and depends on Incumbent's ability to leverage the data.

(b) If RDP is enforced, customers of Incumbent in Stage 1 can port their data to Entrant in Stage 2, and enjoy additional utility at Entrant through inflated tendency to disclose. This is due to the customized services based on prior data, and depends on Entrant's ability to leverage the data.

(c) Entrant's data leverage ability is less than that of Incumbent.

We capture the data leverage ability of the platforms as coefficients for Incumbent and Entrant, denoted as $\alpha_I > 1$ and $\alpha_E \geq 1$, respectively. These coefficients inflate the customers' tendency to disclose to $\alpha_I w$ or $\alpha_E w$, which adds to the total utility customers receive. This assumption applies to personalized products and services that platforms can offer to individuals with a prior interaction with the platform. This is seen in a variety of industries including retail, social networks, communications, online advertising, entertainment, and cloud services. Our model does not capture scenarios where there are externalities in customers' data, i.e. where aggregated data on prior customers can be used to create better services for new customers. It is reasonable to assume that Incumbent is better at leveraging the data created in its own platform, because of its expertise and familiarity with its own operations. This also reflects Entrant's additional effort needed to make use of the customer data at another platform, i.e. $\alpha_E < \alpha_I$.

ASSUMPTION 5 **(Free Porting and No Price Discrimination)**. *(a) Customers incur no cost for data porting.*

(b) Platforms can set only one price in each stage, which applies to customers with or without prior data.

This implies that there are simple and easy-to-use standards available for the data porting. While there are no agreed-upon standards currently in place, there have been some efforts by platforms and privacy advocates to make it easier to port. Some examples include development apps and tools for seamless data porting between platforms (for example between Android and iOS mobile platforms), release of the data in easy-to-port formats such as "json" (Ip 2018) or use of Application Program Interfaces (APIs), and initiatives such as Data Transfer Project and Universal Digital Profile. We do not consider any discount, promotion, or extra cost for customers that have prior data or port their data in Stage 2.

ASSUMPTION 6 **(No Information Asymmetry)**. *All model parameters are known to both platforms.*

With no information asymmetry, we focus on the impacts of enforcing RDP.

ASSUMPTION 7 **(Technical Assumption)**. *We assume that the customer located at location 1 receives a non-positive utility from Incumbent in Stage 1.*

The sufficient condition for this assumption is to have $v \leq 2t$. This simplifies the analysis by limiting the maximum demand in Stage 1 to one unit, but does not impact our results. This also ensures that the Entrant enters the market.

Table 1 summarizes the notation.

Table 1 Notation

Notation	Definition
v	Intrinsic valuation of customers for platform use; $v \in R^+$
t	Customer mismatch cost for platform use based on the Hotelling model; $t \in R^+$
D	Customer data disclosed to platforms through platform use; $D \in R^+$
w	Tendency of customers to use platforms and disclose data; $w \in R^+$
α_I	Coefficient for Incumbent's ability to leverage data; $\alpha_I > 1$
α_E	Coefficient for Entrant's ability to leverage data; $\alpha_E \geq 1$, $\alpha_E < \alpha_I$

3. Stage 1: Incumbent Platform

In Stage 1, Incumbent sets price for its service to maximize its profit. We assume heterogeneous customers according to a Hotelling model, where customers' valuation of the service is horizontally differentiated over the range $[0,1]$, with a total normalized market size of 1. Incumbent locates at 0 and Entrant locates at 1[5]. The base utility of a customer located at $y \in [0,1]$ for Incumbent (the single platform at this stage) is $v - ty$. The total customer utility also depends on how much the customers use the platforms and disclose data, as well as the price for the platform. As per Assumption 3, the utility from using the platform and disclosing D_{I1} unit of data is $D_{I1}[w - D_{I1}]$,

[5] It would be optimal for Incumbent platform to self-locate at $1/2$ in the first stage. For simplicity, we do not consider the location decision, we assume that Incumbent is located at 0. It can be shown that this does not impact our results, because the Entrant does not have access to the market where customers are located to one side of Incumbent, and therefore, that part of the demand is of no interest to our analysis. In this sense, we are not interested in the location problem, but the profit maximization problem where the location is already set.

which is multiplied by the base utility for each customer to derive the total utility for each customer. For a customer that is located at $y \in [0,1]$ on the Hotelling line, the customer utility is derived as:

$$U_{I1}(y) = [v - ty]D_{I1}[w - D_{I1}] - P_{I1}, \tag{1}$$

where $U_{I1}(y)$ is the utility that a customer located at y receives from using Incumbent in Stage 1, and P_{I1} is the price for using Incumbent in Stage 1. According to Assumption 3, each customer optimizes their utility by deciding how much data to disclose. The optimal disclosure for the utility function in (1) for a given disclosure tendency w is $D^*_{I1} = w/2$. Therefore, the utility function of the customers considering the optimal disclosure is derived as:

$$U_{I1}(y) = \frac{[v - ty]w^2}{4} - P_{I1} \tag{2}$$

Customers with $U_{I1}(y) \geq 0$ use the platform in Stage 1. Using the utility function in (2), we can find the location of the customer that is indifferent between using and not using Incumbent in Stage 1 as $\hat{y}_{I1} = [vw^2 - 4P_{I1}]/tw^2$. Customers with $y \leq \hat{y}_{I1}$ use Incumbent and enjoy a positive utility. If $\hat{y}_{I1} = 1$, then market is covered, that is, $q_{I1} = 1$. As per Assumption 7, $\hat{y}_{I1}$ cannot be higher than 1. The demand for Incumbent in Stage 1 is therefore taken as $q_{I1} = \hat{y}_{I1} = [vw^2 - 4P_{I1}]/tw^2$. Incumbent's profit is $\Pi_{I1} = q_{I1}P_{I1}$, which is optimized at the optimal price of $P^*_{I1} = vw^2/8 \geq 0$. The optimal demand and profit are derived as:

$$q^*_{I1} = \frac{v}{2t}, \tag{3}$$

$$\Pi^*_{I1} = \frac{v^2w^2}{16t}. \tag{4}$$

As a participation constraint for Incumbent in Stage 1, we assume $0 \leq \hat{y}^*_{I1}$, which is true given that v and t are positive.

4. Stage 2: Two Platforms

In Stage 2, Entrant, located at 1 on the Hotelling line, enters the market. From Assumption 4, the collected data in Stage 1 can be used to provide better services to those customer that used the platform, i.e. customers with $y \leq q^*_{I1}$. This increases the utility of platform for such customers through increased tendency to disclose (w). We model the additional tendency to disclose as a relative increase according to the (data) leverage coefficient (α_I). The utility that customers enjoy from using Incumbent in Stage 2 is therefore given as:

$$U_{I2}(y) = \begin{cases} [v - ty]D_{I2}[\alpha_I w - D_{I2}] - P_{I2}, & \text{if } y \leq q^*_{I1} \\ \\ [v - ty]D_{I2}[w - D_{I2}] - P_{I2}, & \text{if } y > q^*_{I1} \end{cases}$$

where D_{I2} is the disclosure to Incumbent in Stage 2, and q^*_{I1} is Incumbent's Stage 1 demand given in (3). The optimal amount of disclosure is derived as $D^*_{I2} = \alpha_I w/2$ and $D^*_{I2} = w/2$, for customers that did and did not use Incumbent in Stage 1, respectively. This can be used to derive the utility as:

$$U_{I2}(y) = \begin{cases} \frac{\alpha_I^2 w^2 [v-ty]}{4} - P_{I2}, & \text{if } y \leq q^*_{I1} \\ \\ \frac{w^2 [v-ty]}{4} - P_{I2}, & \text{if } y > q^*_{I1} \end{cases} \tag{5}$$

The utility for Entrant depends on whether Incumbent's Stage 1 customers can port their data to Entrant in Stage 2, i.e. whether data portability is made possible through enforcing the RDP.

4.1. No RDP

If porting is not possible, then the customer utility for the Entrant is similar to Incumbent in Stage 1 (equation 2), but with the platform located at 1 instead of 0. The optimal customer data disclosure in this case is $D^*_E = w/2$, which yields the following utility for customers:

$$U_E^{NP}(y) = \frac{[v - t[1 - y]]w^2}{4} - P_E \tag{6}$$

where $U_E^{NP}(y)$ is the utility for a customer located at y from Entrant in Stage 2 without data porting, and P_E is the price of Entrant in Stage 2. Using the utility equations for Incumbent and Entrant in Stage 2 (equations 5 and 6), we can find the location of the customer that is indifferent between using Incumbent, and Entrant without data porting. The indifferent customer's location depends on comparison of the three utility functions: Incumbent for Incumbent's Stage 1 customers,

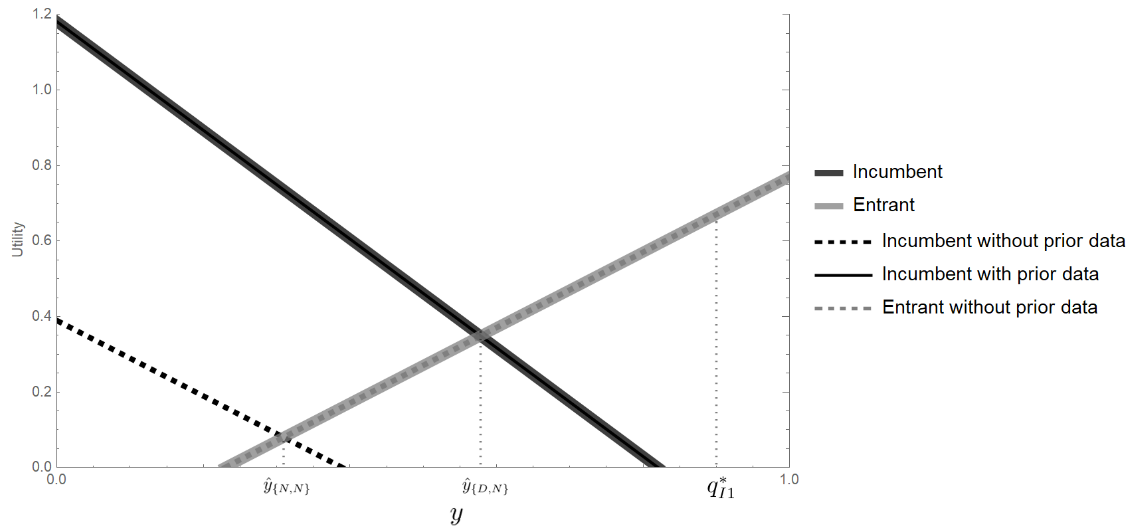


Figure 1 Customer utility for platforms without RDP (Case 1: $w=2$**,** $t=1$**,** $v=1.8$**,** $\alpha_I=1.2$**)**

Incumbent for Incumbent's Stage 1 non-customers, and Entrant without data porting. Figure 1 shows these utility functions for an example.

Depending on location of the indifferent customers between these utilities relative to q^*_{I1}, we can find the different scenarios for the customer that is indifferent between the two platforms. We denote the location of indifferent customer between $U_{I2_{y\leq q^*_{I1}}}(y)$ and $U_E^{NP}(y)$ as $\hat{y}_{\{D,N\}}$; and between $U_{I2_{y>q^*_{I1}}}(y)$ and $U_E^{NP}(y)$ as $\hat{y}_{\{N,N\}}$. In other words, $U_{I2_{y\leq q^*_{I1}}}(\hat{y}_{\{D,N\}}) = U_E^{NP}(\hat{y}_{\{D,N\}})$, and $U_{I2_{y>q^*_{I1}}}(\hat{y}_{\{N,N\}}) = U_E^{NP}(\hat{y}_{\{N,N\}})$. In this notation, the first and second subscript of $\hat{y}$ denote the availability (D) or unavailability (N) of prior customer data for the Incumbent and Entrant, respectively. Because $U_{I2_{y\leq q^*_{I1}}}(y) > U_{I2_{y>q^*_{I1}}}(y)$ due to the added utility from customers' Stage 1 data for customers with $y \leq q^*_{I1}$, we have $\hat{y}_{\{N,N\}} < \hat{y}_{\{D,N\}}$. There are three possible cases:

Case 1: Incumbent's Market Share Shrinks (Customers Switch) If $\hat{y}_{\{D,N\}} \leq q^*_{I1}$ (which implies $\hat{y}_{\{N,N\}} < q^*_{I1}$), then the indifferent customer between the two platforms is located at $\hat{y} = \hat{y}_{\{D,N\}}$. Customers with $y \leq \hat{y}_{\{D,N\}}$, that used Incumbent in Stage 1, choose Incumbent again and benefit from the customized service. Customers with $y > \hat{y}_{\{D,N\}}$, choose Entrant. Incumbent's market share decreases in this case compared to Stage 1. Switching customers are characterized as $\hat{y}_{\{D,N\}} \leq y \leq q^*_{I1}$.

Case 2: No Change in Incumbent's Market Share If $\hat{y}_{\{D,N\}} > q^*_{I1}$ and $\hat{y}_{\{N,N\}} \leq q^*_{I1}$, then the customer that is indifferent between the two platforms is located at $\hat{y} = q^*_{I1} = v/2t$. Customers with $y \leq q^*_{I1}$ choose Incumbent and customers with $y > q^*_{I1}$, choose Entrant. Incumbent's market share is unchanged in this case compared to Stage 1, and Entrant takes the remaining demand. In equilibrium, both platforms set prices so that the customer located at q^*_{I1} has a utility of zero, i.e. $U_{I2_{y \leq q^*_{I1}}}(q^*_{I1}) = U_E(q^*_{I1}) = 0$.

Case 3: Incumbent's Market Share Expands If $\hat{y}_{\{N,N\}} > q^*_{I1}$ (which implies $\hat{y}_{\{D,N\}} > q^*_{I1}$), then the customer that is indifferent between the two platforms is located at $\hat{y}_{\{N,N\}}$. Customers with $y \leq \hat{y}_{\{N,N\}}$ choose Incumbent, and customers with $0 \leq y \leq q^*_{I1}$ benefit from customized service based on their Stage 1 usage. Customers with $y > \hat{y}_{\{N,N\}}$, choose Entrant. Customers with $q^*_{I1} \leq y \leq \hat{y}_{\{N,N\}}$ that did not use Incumbent in Stage 1, use this platform in Stage 2. No Incumbent customer switches to Entrant in this case.

We can combine these three cases to characterize the indifferent customer as follows.

$$\hat{y} = \begin{cases} \hat{y}_{\{D,N\}} = \frac{w^2[t+v[\alpha_I^2-1]]+4[P_E-P_{I2}]}{tw^2[1+\alpha_I^2]} & \text{Case 1: if } \hat{y}_{\{D,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{N,N\}} \leq q^*_{I1} \\ q^*_{I1} = \frac{v}{2t} & \text{Case 2: if } \hat{y}_{\{D,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{N,N\}} \leq q^*_{I1} \\ \hat{y}_{\{N,N\}} = \frac{tw^2+4[P_E-P_{I2}]}{2tw^2} & \text{Case 3: if } \hat{y}_{\{D,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{N,N\}} > q^*_{I1} \end{cases}$$

We return to these conditions after deriving the equilibria. Demands for the two platforms are given as $q_{I2} = \hat{y}$ and $q_E = 1 - \hat{y}$, and the profits are given as $\Pi_{I2} = q_{I2}P_{I2}$ and $\Pi_E = q_E P_E$. We can use these to derive the equilibrium prices. We find only one set of equilibrium prices for each of the three cases as follows:

$$P^*_{I2} = \begin{cases} \frac{w^2[t[2+\alpha_I^2]+v[\alpha_I^2-1]]}{12}, & \text{Case 1: if } \hat{y}_{\{D,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{N,N\}} \leq q^*_{I1} \\ \frac{w^2 v \alpha_I^2}{8}, & \text{Case 2: if } \hat{y}_{\{D,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{N,N\}} \leq q^*_{I1} \\ \frac{tw^2}{4}, & \text{Case 3: if } \hat{y}_{\{D,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{N,N\}} > q^*_{I1} \end{cases} \tag{7}$$

$$P_E^* = \begin{cases} \frac{w^2[t[1+2\alpha_I^2]-v[\alpha_I^2-1]]}{12}, & \text{Case 1: if } \hat{y}_{\{D,N\}} \leq q_{I1}^* \text{ and } \hat{y}_{\{N,N\}} \leq q_{I1}^* \\ \frac{w^2[3v-2t]}{8}, & \text{Case 2: if } \hat{y}_{\{D,N\}} > q_{I1}^* \text{ and } \hat{y}_{\{N,N\}} \leq q_{I1}^* \\ \frac{tw^2}{4}, & \text{Case 3: if } \hat{y}_{\{D,N\}} > q_{I1}^* \text{ and } \hat{y}_{\{N,N\}} > q_{I1}^* \end{cases} \tag{8}$$

We can use the equilibrium prices to derive the conditions for these cases. By substituting the equilibrium prices, the conditions for Case 1 ($\hat{y}_{\{D,N\}} \leq q_{I1}^*$ and $\hat{y}_{\{N,N\}} \leq q_{I1}^*$) can be derived as $[1+\frac{\alpha_I^2-1}{5+\alpha_I^2}]t < v$. However, we can derive from the covered market assumption (Assumption 2) that in this case we need to have $K(\alpha_I)t < v$, where $K(\alpha_I) = 1 + \frac{1+\alpha_I^2+\alpha_I^4}{1+4\alpha_I^2+\alpha_I^4}$. We refer to $K(\alpha_I)$ as K. It can be seen that $1 < K < 2$. Because $1 + \frac{\alpha_I^2-1}{5+\alpha_I^2} < K$, the only condition needed for this case is market coverage. Case 2 ($\hat{y}_{\{D,N\}} > q_{I1}^*$ and $\hat{y}_{\{N,N\}} \leq q_{I1}^*$) is straightforward. The conditions in Case 3 ($\hat{y}_{\{D,N\}} > q_{I1}^*$ and $\hat{y}_{\{N,N\}} > q_{I1}^*$) can be derived in equilibrium as $v \leq t$, however, from the covered market assumption in this case we need to have $3t/2 \leq v$, which is a contradiction. Therefore, there is no equilibrium in this case, i.e. Incumbent's market share cannot expand under market coverage assumption.

The conditions for the remaining two cases (Cases 1 and 2) can be simplified as:

$$\hat{y} = \begin{cases} \frac{t[2+\alpha_I^2]+v[\alpha_I^2-1]}{3t[1+\alpha_I^2]}, & \text{Case 1: if } Kt < v \leq 2t \\ \frac{v}{2t}, & \text{Case 2: if } t \leq v \leq Kt \end{cases} \tag{9}$$

Using the indifferent customer location and equilibrium prices, equilibrium demand and profit are derived as:

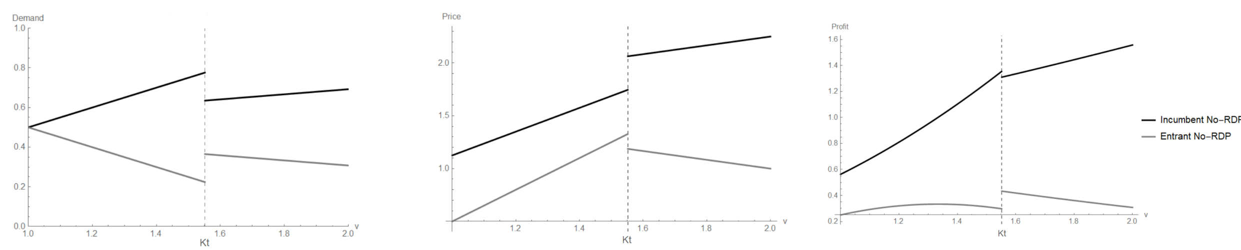
$$q_{I2}^* = \begin{cases} \frac{t[2+\alpha_I^2]+v[\alpha_I^2-1]}{3t[1+\alpha_I^2]}, & \text{Case 1: if } Kt < v \leq 2t \\ \frac{v}{2t}, & \text{Case 2: if } t \leq v \leq Kt \end{cases} \tag{10}$$

$$q_E^* = \begin{cases} \frac{t[1+2\alpha_I^2]-v[\alpha_I^2-1]}{3t[1+\alpha_I^2]}, & \text{Case 1: if } Kt < v \leq 2t \\ 1-\frac{v}{2t}, & \text{Case 2: if } t \leq v \leq Kt \end{cases} \tag{11}$$

$$\Pi_{I2}^* = \begin{cases} \frac{w^2[t[2+\alpha_I^2]+v[\alpha_I^2-1]]^2}{36t[1+\alpha_I^2]}, & \text{Case 1: if } Kt < v \leq 2t \\ \frac{w^2v^2\alpha_I^2}{16t}, & \text{Case 2: if } t \leq v \leq Kt \end{cases} \tag{12}$$

$$\Pi_E^* = \begin{cases} \frac{w^2[t[1+2\alpha_I^2]-v[\alpha_I^2-1]]^2}{36t[1+\alpha_I^2]}, & \text{Case 1: if } Kt < v \leq 2t \\ \frac{w^2[8tv-3v^2-4t^2]}{16t}, & \text{Case 2: if } t \leq v \leq Kt \end{cases} \quad (13)$$

Figure 2 provides the demand, price, and profit of Incumbent and Entrant as intrinsic valuation (v) varies compared to Incumbent's ability to leverage data (α_I). These figures show the Case 1 ($Kt < v \leq 2t$) and Case 2 ($t \leq v \leq Kt$) equilibria and the comparison between Incumbent and Entrant.



(a) Demands (b) Prices (c) Profits

Figure 2 Incumbent versus Entrant without RDP with respect to customer intrinsic valuation ($\alpha_I = 1.5$, $t = 1$, $w = 2$)

Consider the differences between the two platforms in previous equations. Incumbent's demand and profit are always higher than those of Entrant's. This is because Incumbent benefits from aggregated customer data from Stage 1, and because the market is covered. Entrant has no advantage over Incumbent to be able to realize higher demand or profit. Incumbent, however, sets lower prices than Entrant when $\alpha_I < \sqrt{2}$ and $v > 2t/[3-\alpha^2]$. This is due to the price competition that Incumbent initiates in order to prevent losing customers to Entrant when its data leverage ability is poor.

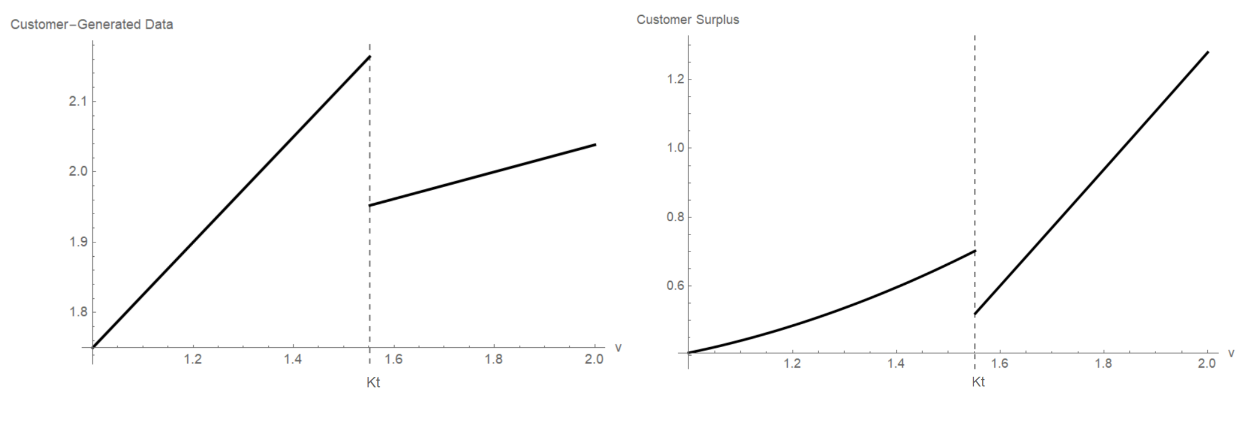
We are also interested in studying the customer usage and data disclosure. Total data generated (or shared) by customers in the case without RDP is derived as:

$$TD = \begin{cases} \int_0^{\hat{y}\{D,N\}} \frac{w\alpha_I}{2} dy + \int_{\hat{y}\{D,N\}}^1 \frac{w}{2} dy = \\ \frac{w[\alpha_I v[\alpha_I^2-1]+t[3+\alpha_I[1+\alpha_I][2+\alpha_I]]]}{6t[1+\alpha_I^2]}, & \text{Case 1: if } Kt < v \le 2t \\ \int_0^{q^*_{I1}} \frac{w\alpha_I}{2} dy + \int_{q^*_{I1}}^1 \frac{w}{2} dy = \\ \frac{w[2t+v\alpha_I]]}{4t}, & \text{Case 2: if } t \le v \le Kt \end{cases} \quad (14)$$

Customer surplus in the case without RDP enforcement is derived as:

$$S = \begin{cases} \int_0^{\hat{y}\{D,N\}} U^*_{I2}(y) dy + \int_{\hat{y}\{D,N\}}^1 U_E^{NP^*}(y) dy = \\ \frac{w^2[v^2[\alpha_I^2-1]^2+4tv[2+5\alpha_I^2+2\alpha_I^4]-t^2[11+23\alpha_I^2+11\alpha_I^4]]}{72t[1+\alpha_I^2]}, & \text{Case 1: if } Kt < v \le 2t \\ \int_0^{q^*_{I1}} U^*_{I2}(y) dy + \int_{q^*_{I1}}^1 U_E^{NP^*}(y) dy = \\ \frac{w^2[4t^2-4tv+v^2[1+\alpha_I^2]]}{32t}, & \text{Case 2: if } t \le v \le Kt \end{cases} \quad (15)$$

Figure 3 provides the total customer-generated data and customer surplus as intrinsic valuation (v) varies compared to Incumbent's ability to leverage data (α_I). These figures show the Case 1 ($Kt < v \le 2t$) and Case 2 ($t \le v \le Kt$) equilibria.



(a) Customer-generated data (b) Customer surplus

Figure 3 Incumbent versus Entrant without RDP with respect to customer intrinsic valuation ($\alpha_I = 1.5$, $t = 1$, $w = 2$)

4.2. RDP

If RDP is enforced, then customers of Incumbent in Stage 1 can port their data to Entrant, given that it increases their utility over the utility they receive from Incumbent. The utility that customers receive from Incumbent in Stage 2 is similar to the case without RDP (5). Entrant, however, can potentially yield higher utility compared to the case without RDP to customers with a preference for Entrant. Only customers of Incumbent in Stage 1 have generated data that they can port to Entrant in Stage 2, i.e., customers with $y \leq q^*_{I1}$, where q^*_{I1} is given in (3). The utility of customers that port their data to Entrant is derived as:

$$U_{E_{y\leq q^*_{I1}}}(y) = [v - t[1-y]]D_E[\alpha_E w - D_E] - P_E,$$

where D_E is the data disclosure amount to Entrant in Stage 2, and α_E is the data leverage coefficient for Entrant. The optimal customer disclosure in this case is $D^*_E = \alpha_E w/2$. Customers who didn't use Incumbent in Stage 1, get the same utility as the case without RDP. Combining the customers that have and have not generated data in Stage 1 yields the following utility function:

$$U_E(y) = \begin{cases} \frac{\alpha_E^2 w^2[v-t[1-y]]}{4} - P_E, & \text{if } y \leq q^*_{I1} \\ \frac{w^2[v-t[1-y]]}{4} - P_E, & \text{if } y > q^*_{I1} \end{cases} \tag{16}$$

Because there is no cost to porting the data, customer utility if data is ported $(y \leq q^*_{I1})$ is higher than if it is not ported $(y > q^*_{I1})$.

Using the utility equations for Incumbent and Entrant in Stage 2, (5) and (16), we can find the location of the customer that is indifferent between using Incumbent and Entrant. The indifferent customer's location depends on comparison of the four utility functions: Incumbent for Incumbent's Stage 1 customers, Incumbent for Incumbent's Stage 1 non-customers, Entrant for Incumbent's Stage 1 customers, and Entrant for Incumbent's Stage 1 non-customers. Figure 4 shows these utility functions for an example.

Depending on the location of indifferent customer relative to q^*_{I1}, we can find the different scenarios for the customer that is indifferent between the two platforms. We denote the location of the indifferent customer between the different utility functions as follows:

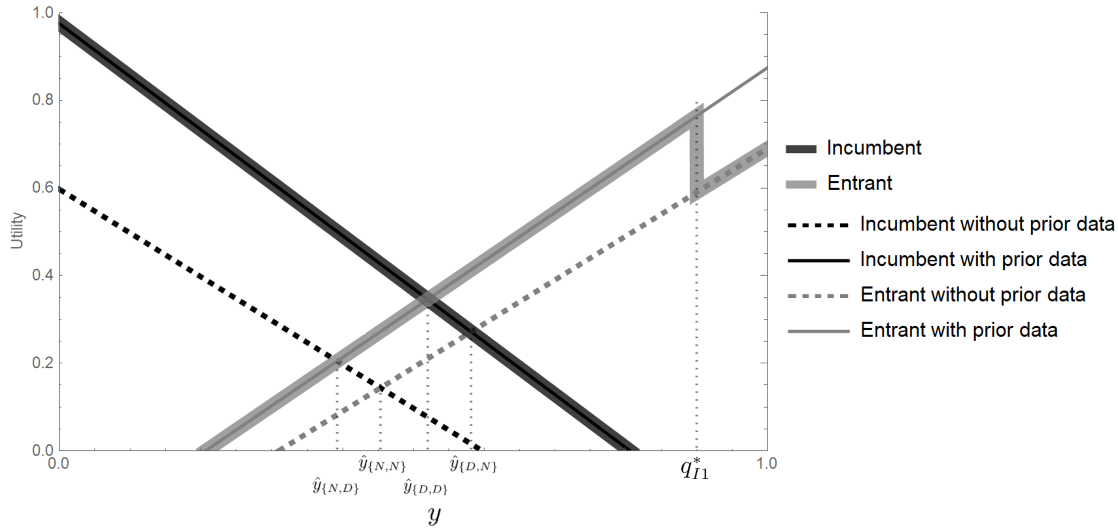


Figure 4 **Customer utility for platforms with RDP (Case 1:** $w = 2$, $t = 1$, $v = 1.8$, $\alpha_I = 1.1$, $\alpha_E = 1.05$**)**

- $\hat{y}_{\{D,D\}}$: between $U_{I2_{y\le q^*_{I1}}}(y)$ and $U_{E_{y\le q^*_{I1}}}(y)$, i.e. $U_{I2_{y\le q^*_{I1}}}(\hat{y}_{\{D,D\}}) = U_{E_{y\le q^*_{I1}}}(\hat{y}_{\{D,D\}})$
- $\hat{y}_{\{D,N\}}$: between $U_{I2_{y\le q^*_{I1}}}(y)$ and $U_{E_{y> q^*_{I1}}}(y)$, i.e $U_{I2_{y\le q^*_{I1}}}(\hat{y}_{\{D,N\}}) = U_{E_{y> q^*_{I1}}}(\hat{y}_{\{D,N\}})$
- $\hat{y}_{\{N,D\}}$: between $U_{I2_{y> q^*_{I1}}}(y)$ and $U_{E_{y\le q^*_{I1}}}(y)$, i.e. $U_{I2_{y> q^*_{I1}}}(\hat{y}_{\{N,D\}}) = U_{E_{y\le q^*_{I1}}}(\hat{y}_{\{N,D\}})$
- $\hat{y}_{\{N,N\}}$: between $U_{I2_{y> q^*_{I1}}}(y)$ and $U_{E_{y> q^*_{I1}}}(y)$, i.e. $U_{I2_{y> q^*_{I1}}}(\hat{y}_{\{N,N\}}) = U_{E_{y> q^*_{I1}}}(\hat{y}_{\{N,N\}})$

Because $U_{I2_{y\le q^*_{I1}}}(y) > U_{I2_{y> q^*_{I1}}}(y)$ and $U_{E_{y\le q^*_{I1}}}(y) > U_{E_{y> q^*_{I1}}}(y)$ due to the added utility from customers' Stage 1 data for those with $y \le q^*_{I1}$, we have $\hat{y}_{\{N,N\}} < \hat{y}_{\{D,N\}}$, $\hat{y}_{\{N,D\}} < \hat{y}_{\{D,D\}}$, $\hat{y}_{\{N,N\}} > \hat{y}_{\{N,D\}}$, and $\hat{y}_{\{D,N\}} > \hat{y}_{\{D,D\}}$. There are four possible cases depending on the location of these intersection points relative to q^*_{I1}.

Case 1: Incumbent's Market Share Shrinks (Customers Switch and Port Data) If $\hat{y}_{\{D,D\}} \le q^*_{I1}$ and $\hat{y}_{\{N,N\}} \le q^*_{I1}$, then the indifferent customer between the two platforms is located at $\hat{y}_{\{D,D\}}$. Customers with $y \le \hat{y}_{\{D,D\}}$, that used Incumbent in Stage 1, choose Incumbent again and benefit from the customized service. Customers with $y > \hat{y}_{\{D,D\}}$, choose Entrant. In this case, Incumbent's market share decreases compared to Stage 1. Switching customers are characterized as $\hat{y}_{\{D,D\}} \le y \le q^*_{I1}$. The market coverage assumption in this case is $U_{E_{y> q^*_{I1}}}(q^*_{I1}) \ge 0$, which implies $U_{I2_{y\le q^*_{I1}}}(\hat{y}_{\{D,D\}}) = U_{E_{y\le q^*_{I1}}}(\hat{y}_{\{D,D\}}) \ge 0$.

Case 2: No Change in Incumbent's Market Share If $\hat{y}_{\{D,D\}} > q^*_{I1}$ (which implies $\hat{y}_{\{D,N\}} > q^*_{I1}$) and $\hat{y}_{\{N,N\}} \le q^*_{I1}$, then the customer that is indifferent between the two platforms is located at $\hat{y} = q^*_{I1} = v/2t$. Customers with $y \leq q^*_{I1}$ choose Incumbent, and customers with $y > q^*_{I1}$ choose Entrant. In this case, Incumbent's market share is unchanged compared to Stage 1, and Entrant takes the remaining demand. In equilibrium, Entrant sets price so that the customer located at q^*_{I1} without data porting has a utility of zero, i.e. $U_{E_{y>q^*_{I1}}}(q^*_{I1}) = 0$, and Incumbent sets price so that at the limit, $\hat{y}_{\{D,D\}} = q^*_{I1}$.

Case 3: Incumbent's Market Share Expands If $\hat{y}_{\{N,N\}} > q^*_{I1}$ (which implies $\hat{y}_{\{D,N\}} > q^*_{I1}$) and $\hat{y}_{\{D,D\}} > q^*_{I1}$, then the customer that is indifferent between the two platforms is located at $\hat{y} = \hat{y}_{\{N,N\}}$. Customers with $y \leq \hat{y}_{\{N,N\}}$ choose Incumbent and customers with $y > \hat{y}_{\{N,N\}}$, choose Entrant. In this case, Incumbent's market share expands compared to Stage 1.

Case 4: Discontinuous Demand for the two Platforms If $\hat{y}_{\{N,N\}} > q^*_{I1}$ (which implies $\hat{y}_{\{D,N\}} > q^*_{I1}$) and $\hat{y}_{\{D,D\}} < q^*_{I1}$, then the market share for the two platforms is discontinuous, and depends on $\hat{y}_{\{D,D\}}$, $\hat{y}_{\{N,N\}}$, and q^*_{I1} as follows. Customers located in $0 \leq y \leq \hat{y}_{\{D,D\}}$ and $q^*_{I1} \leq y \leq \hat{y}_{\{N,N\}}$ use Incumbent, and customers located in $\hat{y}_{\{D,D\}} \leq y \leq q^*_{I1}$ and $\hat{y}_{\{N,N\}} \leq y \leq 1$ use Entrant.

We can combine these cases to characterize the indifferent customer as follows.

$$\hat{y} = \begin{cases} \hat{y}_{\{D,D\}} = \frac{w^2[t\alpha_E^2 + v[\alpha_I^2 - \alpha_E^2]] + 4[P_E - P_{I2}]}{tw^2[\alpha_I^2 + \alpha_E^2]} & \text{Case 1: if } \hat{y}_{\{N,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} \leq q^*_{I1} \\ q^*_{I1} = \frac{v}{2t} & \text{Case 2: if } \hat{y}_{\{N,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} > q^*_{I1} \\ \hat{y}_{\{N,N\}} = \frac{tw^2 + 4[P_E - P_{I2}]}{2tw^2} & \text{Case 3: if } \hat{y}_{\{N,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} > q^*_{I1} \\ \begin{cases} \hat{y}_{\{D,D\}} = \frac{w^2[t\alpha_E^2 + v[\alpha_I^2 - \alpha_E^2]] + 4[P_E - P_{I2}]}{tw^2[\alpha_I^2 + \alpha_E^2]} \\ \hat{y}_{\{N,N\}} = \frac{tw^2 + 4[P_E - P_{I2}]}{2tw^2} \end{cases} & \text{Case 4: if } \hat{y}_{\{N,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} \leq q^*_{I1} \end{cases}$$

We return to these after deriving the equilibria. Demands for the two platforms for cases 1-3 are given as $q_{I2} = \hat{y}$ and $q_E = 1 - \hat{y}$, and for Case 4 is given as $q_{I2} = \hat{y}_{\{D,D\}} + [\hat{y}_{\{N,N\}} - q^*_{I1}]$ and $q_E = [q^*_{I1} - \hat{y}_{\{D,D\}}] + [1 - \hat{y}_{\{N,N\}}]$. The profits are derived as $\Pi_{I2} = q_{I2}P_{I2}$ and $\Pi_E = q_E P_E$. We can use these to derive the equilibrium prices. We find the following set of equilibrium prices for each of the cases:

$$P^*_{I2} = \begin{cases} \frac{w^2[[t+v]\alpha_I^2+[2t-v]\alpha_E^2]}{12} & \text{Case 1: if } \hat{y}_{\{N,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} \leq q^*_{I1} \\ \frac{w^2[2t[\alpha_E^2-1]+v[3+2\alpha_I^2-2\alpha_E^2]]}{8} & \text{Case 2: if } \hat{y}_{\{N,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} > q^*_{I1} \\ \frac{tw^2}{8} & \text{Case 3: if } \hat{y}_{\{N,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} > q^*_{I1} \\ \frac{w^2[\alpha_I^2[3t+v]+\alpha_E^2[5t-3v]]}{12[2+\alpha_I^2+\alpha_E^2]} & \text{Case 4: if } \hat{y}_{\{N,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} \leq q^*_{I1} \end{cases} \tag{17}$$

$$P^*_{E} = \begin{cases} \frac{w^2[[t+v]\alpha_E^2+[2t-v]\alpha_I^2]}{12} & \text{Case 1: if } \hat{y}_{\{N,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} \leq q^*_{I1} \\ \frac{w^2[3v-2t]}{8} & \text{Case 2: if } \hat{y}_{\{N,N\}} \leq q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} > q^*_{I1} \\ \frac{tw^2}{8} & \text{Case 3: if } \hat{y}_{\{N,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} > q^*_{I1} \\ \frac{w^2[\alpha_I^2[3t-v]+\alpha_E^2[t+3v]]}{12[2+\alpha_I^2+\alpha_E^2]} & \text{Case 4: if } \hat{y}_{\{N,N\}} > q^*_{I1} \text{ and } \hat{y}_{\{D,D\}} \leq q^*_{I1} \end{cases} \tag{18}$$

We can use equilibrium prices to derive the conditions for the cases. By substituting the equilibrium prices, the conditions for Case 1 ($\hat{y}_{\{D,D\}} \leq q^*_{I1}$ and $\hat{y}_{\{N,N\}} \leq q^*_{I1}$) can be derived as $[1+\frac{\alpha_I^2-\alpha_E^2}{\alpha_I^2+5\alpha_E^2}]t < v$. However, we can derive from the covered market assumption that in this case, we need $L(\alpha_I, \alpha_E)t < v$, where

$$L(\alpha_I,\alpha_E) = \begin{cases} 1+\frac{\alpha_I^4+\alpha_I^2\alpha_E^2+\alpha_E^2}{\alpha_I^4+4\alpha_I^2\alpha_E^2+\alpha_E^4}, & \text{if } \alpha_I \leq \sqrt{2} \text{ and } \alpha_E \leq \sqrt{5}/2, \\ 1+\frac{2\alpha_I^2+4\alpha_E^2-3}{9+2\alpha_I^2-2\alpha_E^2}, & \text{otherwise.} \end{cases} \tag{19}$$

We refer to $L(\alpha_I, \alpha_E)$ as L. We are interested in cases where $1 < L < 2$. The reason for having two separate conditions for market coverage in (19) is that when $\alpha_I \leq \sqrt{2}$ and $\alpha_E \leq \sqrt{5}/2$, the market is covered if utility functions are positive at $\hat{y}_{\{D,D\}}$. However, if $\alpha_I > \sqrt{2}$ or $\alpha_E > \sqrt{5}/2$, market is covered if the utility for Entrant is positive at q^*_{I1}. Because $1+\frac{\alpha_I^2-\alpha_E^2}{\alpha_I^2+5\alpha_E^2} < L$, the only condition needed for Case 1 is the market coverage.

In equilibrium for Case 2 ($\hat{y}_{\{D,D\}} > q^*_{I1}$ and $\hat{y}_{\{N,N\}} \leq q^*_{I1}$), Entrant sets price so that the customer located at q^*_{I1} without data porting has a utility of zero, i.e. $U_{E_{y>q^*_{I1}}}(q^*_{I1}) = 0$, and Incumbent sets price so that at the limit, $\hat{y}_{\{D,D\}} = q^*_{I1}$. Substituting for the equilibrium prices, the condition for this case is derived as $v < [1+\frac{\alpha_I^2-\alpha_E^2}{3\alpha_E^2-2-\alpha_I^2}]t$. However, because $1+\frac{\alpha_I^2-\alpha_E^2}{3\alpha_E^2-2-\alpha_I^2} > L$, this is a redundant condition.

The conditions in Case 3 ($\hat{y}_{\{N,N\}} > q^*_{I1}$ and $\hat{y}_{\{D,D\}} > q^*_{I1}$) can be derived as $v \leq t$, however, from the covered market assumption in this case we need to have $3t/2 \leq v$, which is a contradiction, thus this case does not occur, i.e. Incumbent's market share cannot expand. Using the equilibrium prices for Case 4 ($\hat{y}_{\{N,N\}} > q^*_{I1}$ and $\hat{y}_{\{D,D\}} < q^*_{I1}$), we find the condition for this case to be $v \leq [1 - \frac{2\alpha_I^2 - 2\alpha_E^2}{6+5\alpha_I^2-3\alpha_E^2}]t$ which contradict the market coverage condition of $v \geq [1 + \frac{\alpha_I^2+\alpha_E^2-2}{4+2\alpha_I^2+2\alpha_E^2}]t$. Therefore, there is no equilibrium for this case either.

The conditions for the remaining two cases (Case 1 and 2) can be written as:

$$\hat{y} = \begin{cases} \frac{[t+v]\alpha_I^2+[2t-v]\alpha_E^2}{3t[\alpha_I^2+\alpha_E^2]}, & \text{Case 1: if } Lt < v \leq 2t \\ \frac{v}{2t}, & \text{Case 2: if } t \leq v \leq Lt \end{cases} \tag{20}$$

Using the indifferent customer's location and equilibrium prices, the Equilibrium demand and profit are derived as:

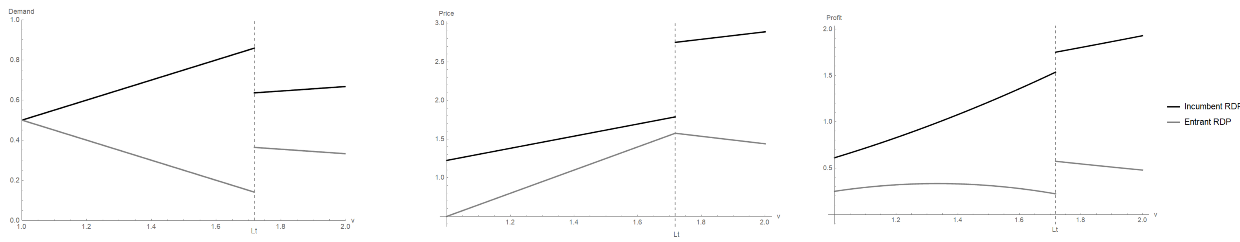
$$q^*_{I2} = \begin{cases} \frac{[t+v]\alpha_I^2+[2t-v]\alpha_E^2}{3t[\alpha_I^2+\alpha_E^2]}, & \text{Case 1: if } Lt < v \leq 2t \\ \frac{v}{2t}, & \text{Case 2: if } t \leq v \leq Lt \end{cases} \tag{21}$$

$$q^*_{E} = \begin{cases} \frac{[t+v]\alpha_I^2+[2t-v]\alpha_I^2}{3t[\alpha_E^2+\alpha_E^2]}, & \text{Case 1: if } Lt < v \leq 2t \\ 1-\frac{v}{2t}, & \text{Case 2: if } t \leq v \leq Lt \end{cases} \tag{22}$$

$$\Pi^*_{I2} = \begin{cases} \frac{w^2[[t+v]\alpha_I^2+[t+v]\alpha_E^2]^2}{36t[\alpha_I^2+\alpha_E^2]}, & \text{Case 1: if } Lt < v \leq 2t \\ \frac{w^2 v[v[3+\alpha_I^2-3\alpha_E^2]+2t[\alpha_E^2-1]]}{16t}, & \text{Case 2: if } t \leq v \leq Lt \end{cases} \tag{23}$$

$$\Pi^*_{E} = \begin{cases} \frac{w^2[[t+v]\alpha_E^2+[t+v]\alpha_I^2]^2}{36t[\alpha_I^2+\alpha_E^2]}, & \text{Case 1: if } Lt < v \leq 2t \\ \frac{w^2[8tv-3v^2-4t^2]}{16t}, & \text{Case 2: if } t \leq v \leq Lt \end{cases} \tag{24}$$

Figure 5 provides the demand, price, and profit of Incumbent and Entrant as intrinsic valuation (v) varies compared to the platforms' ability to leverage data (α_I and α_E). These figures show the Case 1 ($Lt < v \leq 2t$) and Case 2 ($t \leq v \leq Lt$) equilibria and the comparison between Incumbent and Entrant.



(a) Demands (b) Prices (c) Profits

Figure 5 **Incumbent versus Entrant with RDP with respect to customer intrinsic valuation (**$\alpha_I = 1.7$, $\alpha_E = 1.2$, $t = 1$, $w = 2$**)**

Consider the differences between the two platforms in previous equations. Similar to the No-RDP scenario, Incumbent's demand and profit are always higher than those of Entrant's. Incumbent, sets lower prices than Entrant when $\alpha_I < \sqrt{2}$, $\alpha_E < \sqrt{2}$ and $v > 2t[\alpha_E - 1]/[3\alpha_E^2 - \alpha^2 - 2]$. This is due to the price competition that Incumbent initiates in order to prevent losing customers to Entrant when the data leverage abilities are poor.

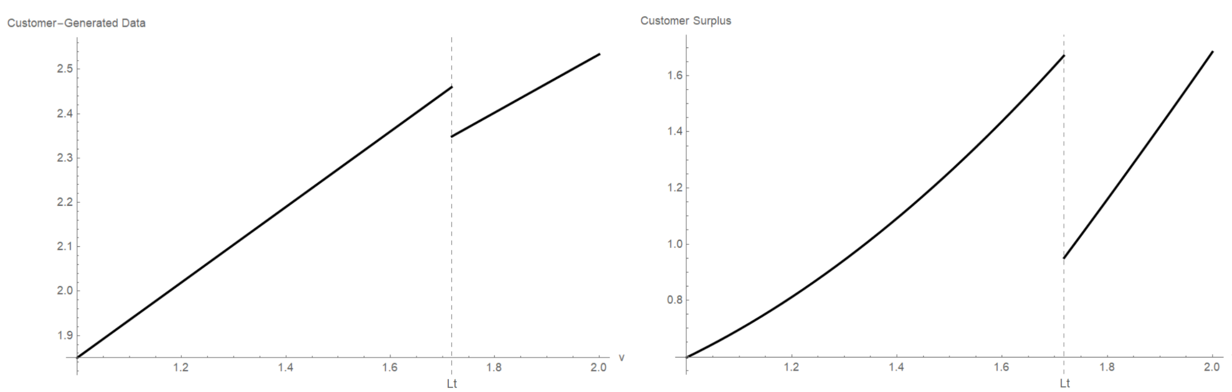
Total data generated by customers in the case with RDP enforcement is derived as:

$$TD = \begin{cases} \int_0^{\hat{y}\{D,D\}} \frac{w\alpha_I}{2} dy + \int_{\hat{y}_{\{D,D\}}}^{q^*_{I1}} \frac{w\alpha_E}{2} dy + \int_{q^*_{I1}}^{1} \frac{w}{2} dy = \\ \frac{w[2\alpha_I^2[v\alpha_I + t[3+\alpha_I]] + [5v-4t]\alpha_E^2 - [2t-v]\alpha_I^2\alpha_E}{12t[\alpha_I^2+\alpha_E^2]} + \\ \frac{w[2\alpha_E^2[t[3+2\alpha_I] - v\alpha_I]]}{12t[\alpha_I^2+\alpha_E^2]}, & \text{Case 1: if } Lt < v \leq 2t \\ \int_0^{q^*_{I1}} \frac{w\alpha_I}{2} dy + \int_{q^*_{I1}}^{1} \frac{w}{2} dy = \\ \frac{w[2t+v\alpha_I]]}{4t}, & \text{Case 2: if } t \leq v \leq Lt \end{cases} \quad (25)$$

Customer surplus in the case with RDP enforcement is derived as:

$$S = \begin{cases} \int_0^{\hat{y}_{\{D,D\}}} U_{I2}^*(y)dy + \int_{\hat{y}_{\{D,D\}}}^1 U_E^*(y)dy = \\ \frac{w^2[v^2[4\alpha_I^4 - 45\alpha_I^2 + 49\alpha_E^2 + \alpha_I[37\alpha_E^2 - 45]}{288t[\alpha_I^2 + \alpha_E^2]} + \\ \frac{w^2[4tv[8\alpha_I^4 + 27\alpha_I^2 - 19\alpha_E^4 + \alpha_I[27 - 7\alpha_E^2]}{288t[\alpha_I^2 + \alpha_E^2]} - \\ \frac{w^2[4t^2[11\alpha_I^4 + 9\alpha_E^2 + 2\alpha_E^4 + \alpha_I[9 + 14\alpha_E^2]]}{288t[\alpha_I^2 + \alpha_E^2]}, & \text{Case 1: if } Lt < v \leq 2t \\ \int_0^{q_{I1}^*} U_{I2}^*(y)dy + \int_{q_{I1}^*}^1 U_E^*(y)dy = \\ \frac{w^2[4t^2 - 4tv\alpha_E^2 + v^2[\alpha_I^2 + 6\alpha_E^2 - 5]]}{32t}, & \text{Case 2: if } t \leq v \leq Lt \end{cases} \tag{26}$$

Figure 6 provides the total customer-generated data and customer surplus as intrinsic valuation (v) varies compared to Incumbent's ability to leverage data (α_I). These figures show the Case 1 ($Lt < v \leq 2t$) and Case 2 ($t \leq v \leq Lt$) equilibria.



(a) Customer-generated data

(b) Customer surplus

Figure 6 **Incumbent versus Entrant with RDP with respect to customer intrinsic valuation (**$\alpha_I = 1.7$**,** $\alpha_E = 1.2$**,** $t = 1$**,** $w = 2$**)**

5. Results

There are different regions to analyze the impact of RDP on platforms, and these regions depend on the comparison of $K(\alpha_I)$ and $L(\alpha_I, \alpha_E)$. We consider two possibilities: $K < L$ and $L < K$. Figure 7 shows the comparison of K and L for different values of α_I and α_E. It can be seen that $L < K$ only when α_E is small, that is, the ability of Entrant to leverage data is relatively poor. This can be the case if the type of data or the technology is in a way that hinders effective leveraging of the data if it is ported to another platform. In other words, the data is incompatible for porting. We refer to this scenario $(L < K)$ as *incompatible porting*, and to the scenario where Entrant's data leverage coefficient is high $(K < L)$ as *compatible porting*.

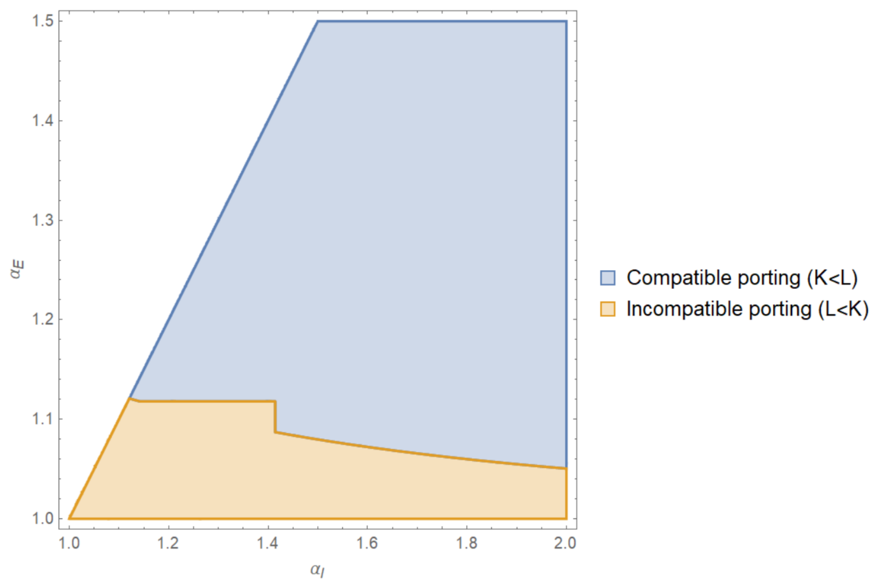
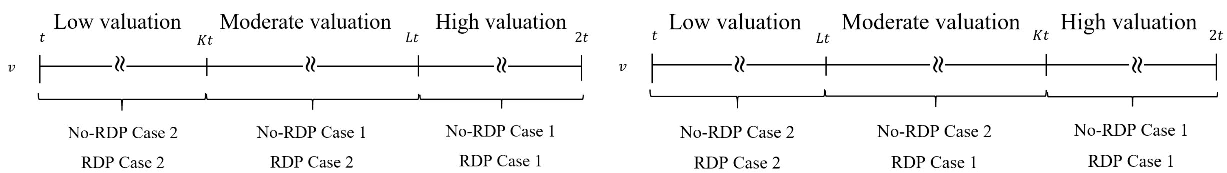


Figure 7 Compatible and incompatible porting regions with respect to the platforms' data leverage coefficients

If porting is compatible $(K < L)$, comparing the conditions for scenarios without RDP (9) to with RDP (20), there exist three regions. If customer valuation is high compared to the mismatch cost and leverage coefficients $(Lt < v \leq 2t)$, then the Case 1 equilibrium for both without and with RDP applies. If valuation is moderate $(Kt < v \leq Lt)$, then the Case 1 equilibrium applies for without RDP and Case 2 equilibrium applies for with RDP. If valuation is low $(t \leq v \leq Kt)$, then the Case 2 equilibrium for both without and with RDP applies.

If porting is incompatible $(L < K)$, the regions change as follows. If customer valuation is high compared to the mismatch cost and leverage coefficients $(Kt < v \leq 2t)$, then the Case 1 equilibrium for both without and with RDP applies. If valuation is moderate $(Lt < v \leq Kt)$, then the Case 2 equilibrium applies for without RDP and Case 1 equilibrium applies for with RDP. If valuation is low $(t \leq v \leq Lt)$, then the Case 2 equilibrium for both without and with RDP applies.

It can be seen that depending on porting compatibility, enforcing RDP can drive platforms to different equilibria, and this determines the impact of RDP on platforms. These scenarios are summarized in Figure 8.



(a) Compatible porting ($K < L$). (b) Incompatible porting ($L < K$).

Figure 8 Equilibrium regions for analyzing the impact of RDP.

High and low valuation equilibria are the same for compatible and incompatible porting scenarios, but are different for the moderate valuation. Therefore, we need to analyze the impact of RDP in four regions: high valuation, moderate valuation and compatible porting ($K < L$), moderate valuation and incompatible porting ($L < K$), and low valuation. While we refer to these scenarios in terms of valuation, it is important to note that valuation is in comparison to mismatch cost, and more importantly, to data leverage coefficients. We analyze the impact of RDP on demands, prices, profits, customer-generated data, and customer surplus in each of these regions.

Lemma 1 **(Demand)**. *Enforcing RDP in Stage 2:*

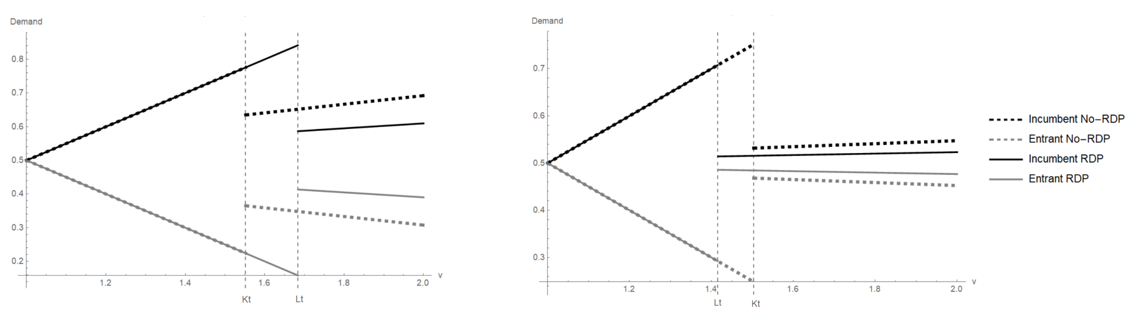
(a) Decreases Incumbent's demand if valuation is high; increases it if valuation is moderate and porting is compatible; decreases it if valuation is moderate and porting is incompatible; and does not change it if valuation is low.

(b) Increases Entrant's demand if valuation is high; decreases it if valuation is moderate and porting is compatible; increases it if valuation is moderate and porting is incompatible; and does not change it if valuation is low.

(c) Increases the number of customers who switch if valuation is high; decreases it if valuation is moderate and porting is compatible; increases it if valuation is moderate and porting is incompatible; and does not change it if valuation is low.

Proof: Comparing the Case 1 and Case 2 demands for No-RDP (10), (11) to RDP (21), (22) it can be seen that (a) Incumbent's RDP demand is lower than its No-RDP demand if $Lt < v \leq 2t$ (where $K < L$) or $Kt < v \leq 2t$ (where $L < K$); is higher if $K < L$ and $Kt < v \leq Lt$; is lower if $L < K$ and $Lt < v \leq Kt$; and is unchanged if $t < v \leq Kt$ (where $K < L$) or $t < v \leq Lt$ (where $L < K$). (b) Entrant's RDP demand behaves inversely to Incumbent's demand. (c) Customers only switch in Case 1 for both No-RDP and RDP scenarios, and the number of customers that switch is derived as $\int_{\hat{y}_{D,N}}^{q_{I1}^*} dy$ and $\int_{\hat{y}_{D,D}}^{q_{I1}^*} dy$ for no-RDP and RDP, respectively. Comparing these for the four regions is straightforward. When market is covered, the number of switching customers follow Entrant's demand. Q.E.D.

Figure 9 provides the impact of RDP on demands for different valuations, and Figure 10 provides the impact of RDP on customer switching.

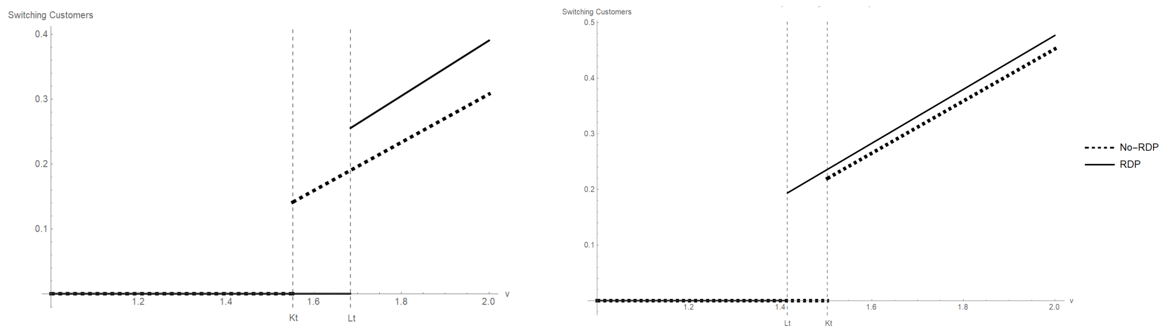


(a) Compatible porting ($\alpha_I = 1.5$, $\alpha_E = 1.2$) (b) Incompatible porting ($\alpha_I = 1.1$, $\alpha_E = 1.05$)

Figure 9 Impact of RDP on platform demands ($t = 1$)

If valuation is high, then introducing RDP decreases the demand for Incumbent to the benefit of Entrant. In high valuation setting, platforms compete over Incumbent's Stage 1 (prior) customers, even if Entrant does not have access to their Stage 1 data. Enforcing RDP provides a boost to the utility that customers get from Entrant, therefore improving its demand.

If valuation is moderate and porting is compatible, interestingly, introducing RDP increases the demand for Incumbent at the expense of Entrant's demand. The reason for this is that if valuation is moderate, while at equilibrium for the case without RDP, Incumbent's demand shrinks compared



(a) Compatible porting ($\alpha_I = 1.5$, $\alpha_E = 1.2$) (b) Incompatible porting ($\alpha_I = 1.1$, $\alpha_E = 1.05$)

Figure 10 Impact of RDP on customer switching ($t = 1$)

to Stage 1, in the case with RDP, platforms divide the market at Incumbent's Stage 1 demand. In this case, if RDP is enforced, due to the reduced price for Incumbent, it is more profitable for Entrant to increase its price (as we see in Lemma 2) and focus on the customers that did not use Incumbent in Stage 1, rather than to compete with Incumbent over its prior customers. In this case, RDP forces firms to switch to limit pricing (Case 2). This change of strategy brings in an interesting effect where if RDP is enforced, it reduces the entrant's demand to the benefit of Incumbent platform's demand. However, if valuation is moderate and porting is incompatible, then the impact of RDP is similar to the case of high valuation, and it increases the demand for Entrant. When porting is incompatible, Incumbent does not need to initiate a price war. This is somewhat counterintuitive where enforcing RDP in moderate valuation setting helps the entrant platform only if it has poor data leverage ability, due to the reduction in price competition in this case.

If valuation is low (while the market is covered), then in both cases with and without RDP, platforms divide the market at Incumbent's Stage 1 demand, and therefore enforcing RDP has no impact. Low valuation implies that Incumbent is really good at leveraging data, which creates a big gap in the utility functions of customers that did and did not use Incumbent in Stage 1. It is therefore not reasonable for the two platforms to compete over Incumbent's prior customers.

Lemma 2 **(Price)**. *Enforcing RDP in Stage 2:*

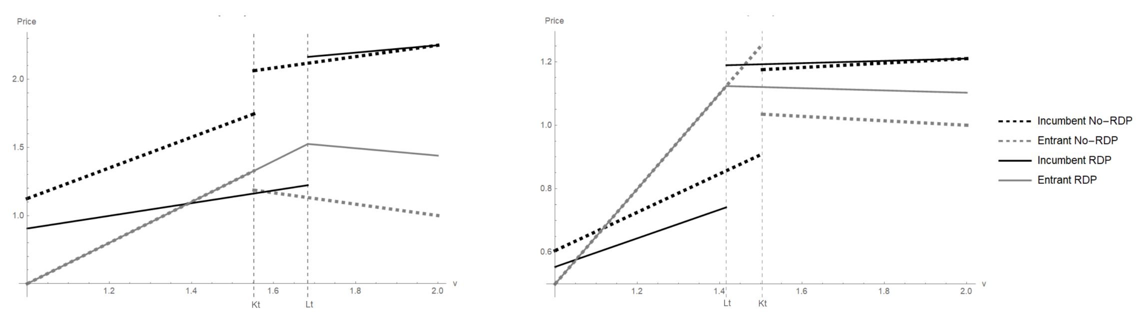
(a) Increases Incumbent's price if valuation is high or if valuation is moderate and porting is incom-

patible; and decreases it if valuation is moderate and porting is compatible or if valuation is low.

(b) Increases Entrant's price if valuation is high or if valuation is moderate and porting is compatible; decreases it if valuation is moderate and porting is incompatible; and does not change it if valuation is low.

Proof: Comparing the Case 1 and Case 2 prices for No-RDP (7), (8) to RDP (17), (18) it can be seen that (a) Incumbent's RDP price is higher than its No-RDP price if $Lt < v \leq 2t$ (where $K < L$) or $Kt < v \leq 2t$ (where $L < K$); is lower if $K < L$ and $Kt < v \leq Lt$; is higher if $Lt < v \leq Kt$ and $L < K$; and is lower if $t < v \leq Kt$ (where $K < L$) or $t < v \leq Lt$ (where $L < K$). (b) Entrant's RDP price is higher than its No-RDP price if $Lt < v \leq 2t$ (where $K < L$) or $Kt < v \leq 2t$ (where $L < K$); is higher if $K < L$ and $Kt < v \leq Lt$; is lower if $Lt < v \leq Kt$ and $L < K$; and is unchanged if $t < v \leq Kt$ (where $K < L$) or $t < v \leq Lt$ (where $L < K$). Q.E.D.

Figure 11 provides the impact of RDP on demands for different valuations.



(a) Compatible porting ($\alpha_I = 1.5$, $\alpha_E = 1.2$) (b) Incompatible porting ($\alpha_I = 1.1$, $\alpha_E = 1.05$)

Figure 11 Impact of RDP on platform prices ($t = 1$, $w = 2$)

In a market with high valuation, if RDP is enforced, Entrant is able to gain some of the market share from Incumbent due to the added utility from their prior data. In this case, it is profitable for Incumbent to specialize and focus on customers closer to it (y close to zero), while increasing the price. Entrant too can increase its price because of the additional utility it can provide due to prior data. This added price for Entrant is spread over all of its customers, whether or not they have prior data.

If valuation is moderate and porting is compatible, if RDP is enforced, Incumbent reduces its price and competes with Entrant to maintain its Stage 1 demand. This is because if valuation is moderate compared to the data leverage coefficient, Incumbent can better utilize the gathered data compared to the case where valuation is high. However, if valuation is moderate and porting is incompatible, then Incumbent is able to increase its price without being worried about competition from Entrant, and Entrant reduces its price.

If valuation is low, then Incumbent initiates a price war to divide the demand to its Stage 1 level for Incumbent. In this case, Incumbent is good at leveraging the data, and this creates a natural dividing point for market share and competition, which is not impacted by RDP.

LEMMA 3 **(Profit)**. *Enforcing RDP in Stage 2:*

(a) Decreases Incumbent's profit.

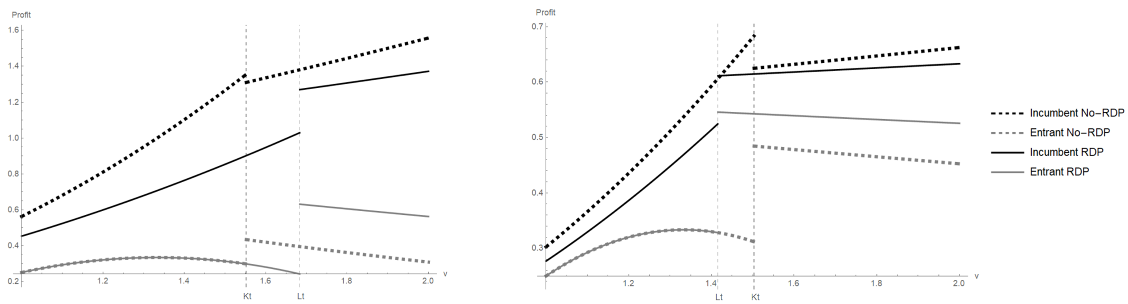
(b) Increases Entrant's profit if valuation is high or if valuation is moderate and porting is incompatible; decreases it if valuation is moderate and porting is compatible; and does not change it if valuation is low.

Proof: Comparing the Case 1 and Case 2 profits for No-RDP (12), (13) to RDP (23), (24) it can be seen that (a) Incumbent's RDP profit is always lower than its No-RDP profit. (b) Entrant's RDP profit is higher than its No-RDP profit if $Lt < v \leq 2t$ (where $K < L$) or $Kt < v \leq 2t$ (where $L < K$); is lower if $K < L$ and $Kt < v \leq Lt$; is higher if $Lt < v \leq Kt$ and $L < K$; and is unchanged if $t < v \leq Kt$ (where $K < L$) or $t < v \leq Lt$ (where $L < K$). Q.E.D.

Figure 12 provides the impact of RDP on profits for different valuations.

This implies that if valuation is high compared to data leverage coefficients, then it is beneficial for Entrant if RDP is implemented. In this case, Entrant's price and demand both increase. While Incumbent increases its price, its effect on profit is smaller than the reduced demand, and therefore, Incumbent's profit decreases.

Interestingly, if valuation is moderate and porting is compatible, then RDP negatively impacts both platforms. For Entrant, this is due to the demand loss, and for Incumbent this is due to



(a) Compatible porting ($\alpha_I = 1.5$, $\alpha_E = 1.2$) (b) Incompatible porting ($\alpha_I = 1.1$, $\alpha_E = 1.05$)

Figure 12 Impact of RDP on platform profits ($t = 1$, $w = 2$)

reduced price. In such a scenario, RDP is bad for both platforms. However, if valuation is moderate and porting is incompatible, RDP negatively impacts Incumbent, but positively impacts Entrant. For Incumbent, this is due to the reduced demand, which has a bigger impact than the increased price. For Entrant, while the price is lower, additional demand more than compensates and profit increases.

If valuation is low, then the Entrant's profit does not change with RDP, as both price and demand are unchanged. However, RDP forces Incumbent to reduce its price due to Entrant's added utility to customers, which reduces its profit.

While our model does not consider entry conditions, we can provide insights on the impact of RDP on Entrant and competition. Table 2 summarizes the effect of RDP on demands, prices, and profits in Lemmas 1, 2, and 3. If valuation is high, or if valuation is moderate and porting is incompatible, then RDP can help increase the demand and profit for Entrant. However, if valuation is moderate and porting is compatible, or if valuation is low, then enforcing RDP results in similar or lower demand and profits for Entrant. This implies that if the goal is to ease competition, regulators should implement RDP only if valuation is high, or the entrant platform's data leverage is poor (incompatible porting). If valuation is high and porting is compatible, enforcing RDP forces Incumbent to reduce its prices and make it harder for Entrant to compete.

Figures 13 and 14 provide the impact of RDP on platforms for two scenarios, one with compatible porting ($K < L$) and one with incompatible porting ($L < K$). It can be seen that as the platforms'

Table 2 Impact of RDP on demand, prices, and profits

	Compatible ($K < L$)	$Lt < v \leq 2t$	$Kt < v \leq Lt$	-	$t \leq v \leq Kt$
	Incompatible ($L < K$)	$Kt < v \leq 2t$	-	$Lt < v \leq Kt$	$t \leq v \leq Lt$
	No-RDP Equilibrium	Case 1	Case 1	Case 2	Case 2
	RDP Equilibrium	Case 1	Case 2	Case 1	Case 2
Demand	Incumbent	$-$	$+$	$-$	0
	Entrant	$+$	$-$	$+$	0
	Switching Customers	$+$	$-$	$+$	0
Price	Incumbent	$+$	$-$	$+$	$-$
	Entrant	$+$	$+$	$-$	0
Profit	Incumbent	$-$	$-$	$-$	$-$
	Entrant	$+$	$-$	$+$	0

data leverage abilities increase, the regions of valuation where enforcing RDP is beneficial to competition, shrinks.

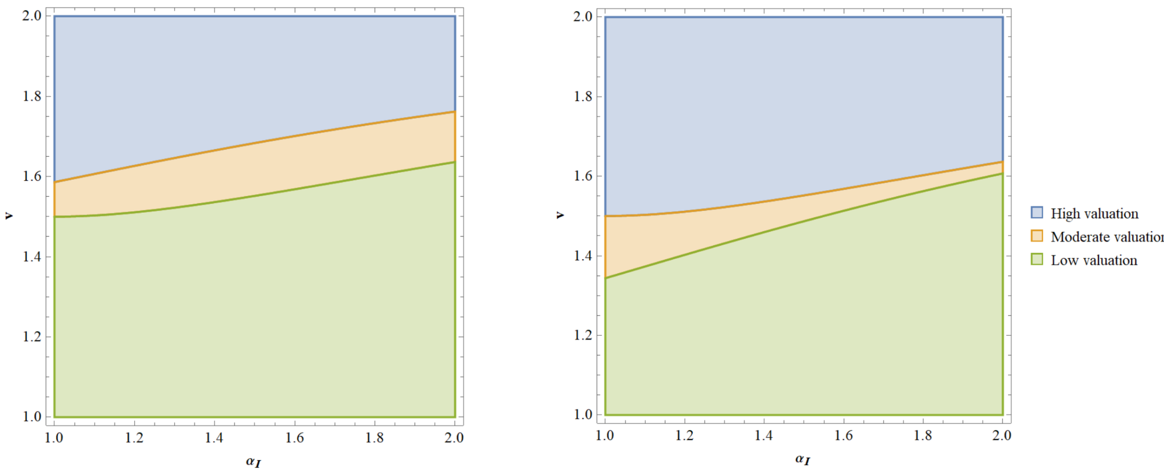
Next, we analyze the impact of RDP on the amount of data generated and customer surplus.

LEMMA 4 **(Total Customer Data and Surplus)**. *Enforcing RDP in Stage 2:*

(a) Increases total customer-generated data if valuation is high, or if valuation is moderate and porting is compatible; decreases it if valuation is moderate and porting is incompatible; and does not change it if valuation is low.

(b) Increases customer surplus if valuation is moderate and porting is compatible, or if valuation is low; and decreases it if valuation is moderate and porting is incompatible. If valuation is high, customer surplus may increase or decrease.

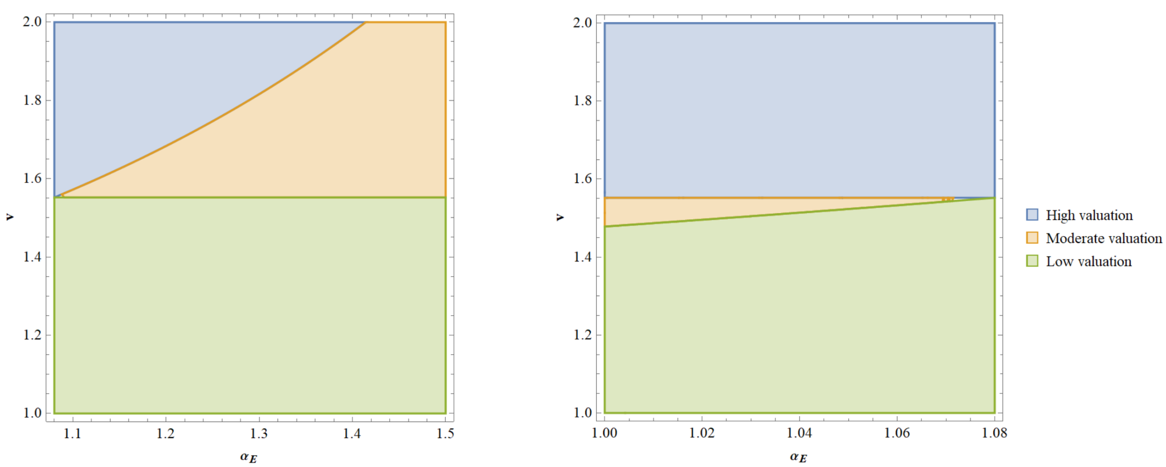
Proof: (a) Comparing the Case 1 and Case 2 customer generated data for No-RDP (14) to RDP (25), it can be seen that the RDP data generated is higher than that of No-RDP if $Lt < v \leq 2t$ (where $K < L$) or $Kt < v \leq 2t$ (where $L < K$); is higher if $K < L$ and $Kt < v \leq Lt$; is lower if



(a) Compatible Porting ($\alpha_E = 1.2$): RDP is beneficial to competition in high valuation region only.

(b) Incompatible Porting ($\alpha_E = 1.01$): RDP is beneficial to competition in high and moderate valuation regions.

Figure 13 Impact of RDP on competition with respect to Incumbent's data leverage coefficient and valuation ($t = 1$).



(a) Compatible porting ($\alpha_E > 0.08$): RDP is beneficial to competition in high valuation region only.

(b) Incompatible porting ($\alpha_E < 0.08$): RDP is beneficial to competition in high and moderate valuation regions.

Figure 14 Impact of RDP on competition with respect to Entrant's data leverage coefficient and valuation ($t = 1$, $\alpha_I = 1.5$).

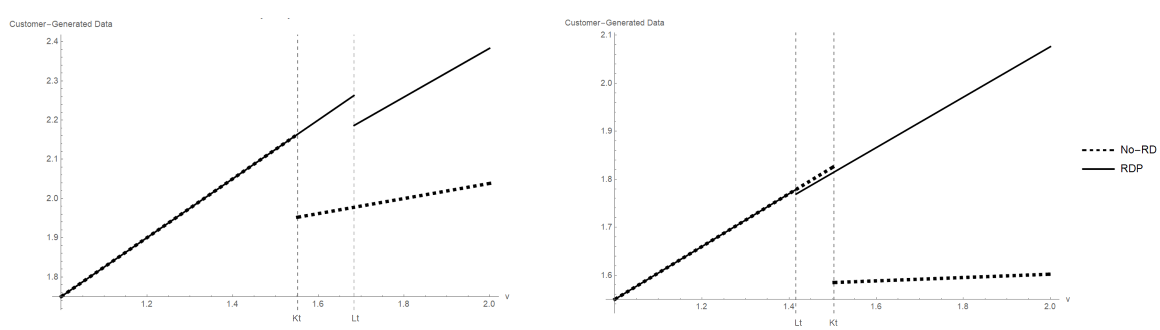
$Lt < v \leq Kt$ and $L < K$; and is unchanged if $t < v \leq Kt$ (where $K < L$) or $t < v \leq Lt$ (where $L < K$). (b) Comparing the Case 1 and Case 2 customer surpluses for No-RDP (15) to RDP (26), it can be seen that the customer surplus for RDP may be higher or lower than that of No-RDP if $Lt < v \leq 2t$ (where $K < L$) or $Kt < v \leq 2t$ (where $L < K$); is higher if $K < L$ and $Kt < v \leq Lt$; is lower if $Lt < v \leq Kt$ and $L < K$; and is higher if $t < v \leq Kt$ (where $K < L$) or $t < v \leq Lt$ (where $L < K$). Q.E.D.

Table 3 provides the summary of Lemma 4.

Table 3 Impact of RDP on customers

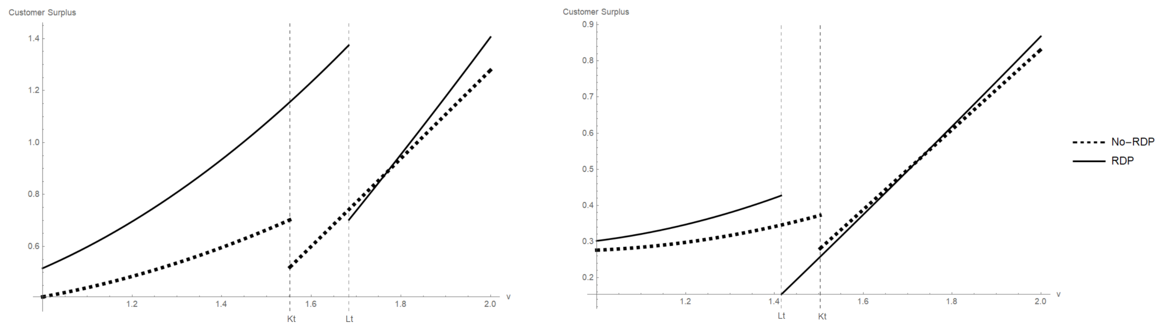
Compatible ($K < L$)	$Lt < v \leq 2t$	$Kt < v \leq Lt$	-	$t \leq v \leq Kt$
Incompatible ($L < K$)	$Kt < v \leq 2t$	-	$Lt < v \leq Kt$	$t \leq v \leq Lt$
No-RDP Equilibrium	Case 1	Case 1	Case 2	Case 2
RDP Equilibrium	Case 1	Case 2	Case 1	Case 2
Total Data	+	+	−	0
Total Customer Surplus	+/−	+	−	+

Figure 15 and 16 provide the impact of RDP on customer-generated data and customer surplus for different valuations.



(a) Compatible porting ($\alpha_I = 1.5$, $\alpha_E = 1.2$) (b) Incompatible porting ($\alpha_I = 1.1$, $\alpha_E = 1.05$)

Figure 15 Impact of RDP on customer-generated data ($t = 1$, $w = 2$)



(a) Compatible porting ($\alpha_I = 1.5$, $\alpha_E = 1.2$) (b) Incompatible porting ($\alpha_I = 1.1$, $\alpha_E = 1.05$)

Figure 16 **Impact of RDP on customer surplus (**$t = 1$**,** $w = 2$**)**

RDP increases the amount of data generated in high valuation, and moderate valuation with compatible porting. This is intuitive, because in our setting where the market is covered, introduction of RDP serves to improve the amount of data generated through the data leverage coefficient. However, if valuation is moderate and porting is incompatible, then due to the reduction of Incumbent' demand, which is not compensated by Entrant due to its low data leverage ability; the total customer-generated data (and customers usage of the platforms) will be reduced.

Total customer surplus in our model captures the impact of prices and data disclosure. The more customers use the platforms and generate data, the more utility they enjoy. In some scenarios, enforcing RDP increases customer surplus, due to the increased utility that customers get from the addition data disclosure, which has a more powerful effect than increased prices. If valuation is high, then enforcing RDP may increase (at higher valuations) or decrease (at lower valuations) customer surplus due to increase in the intensity of use and data generated by customers, which countered by increased prices. If valuation is moderate and porting is compatible, then not only customers enjoy higher intensity of use and higher data generated, but also lower prices for Incumbent. However, if valuation is moderate and porting is incompatible, the intensity of use and data generation is reduced, and the effect of this on surplus is larger than reduction of prices for Entrant. If valuation is small, then enforcing RDP increases surplus due to the decreased price of Incumbent, while the demands and customer-generated data remain the same. These findings are important to regulators with the goal of improving customer surplus. This implies that not only enforcing RDP may not be the best strategy to foster competition, but it may also be bad for customers in certain situations with high and moderate customer valuation.

6. Conclusion and Future Work

Online platforms have been able to leverage the customer-generated data to provide services that are customized to customers' needs. It may be difficult for new entrants to provide equivalent services without access to customer data. The lack of access to customer data acts as a barrier to entry for new entrants. The RDP has been proposed as a solution to this problem, which forces platforms to allow customers to download their data and be able to port it to other platforms. In this paper, we study the impact of RDP on platform demand, pricing, and profits, as well as on customer-generated data and customer surplus. We find that enforcing RDP in some scenarios eases competition by improving the demand and profit for the entrant platform. However, if customer valuation is not high compared to the effectiveness of platforms in leveraging the data and the entrant is competent in leveraging the data, it may in fact be detrimental to competition. In such cases, the Incumbent reduces its price if RDP is enforced, which makes it harder for the entrant platform to compete. This reduces the demand and profit for the entrant compared to the case without RDP. Moreover, if the entrant platform's ability to leverage the data is poor, enforcing RDP may reduce the customer's total usage of the platforms and total customer surplus. These findings have important implications for regulators such as the EU, which seems to assume that data portability is always beneficial for competition and/or customers. Our work shows that this is not always the case.

We provide some interesting insights in this paper from our basic stylized economic model. However, we believe there is much left to be done. We recognize several avenues for extending the current analysis to better understand the impact of policies such as RDP on platforms and competition. First, our model assumes that the market is covered in Stage 2. While this provides a basic model, relaxing this assumption can provide more general insights. Without market coverage assumption, it is possible to study conditions for market entry, which are crucial for improving regulatory decisions. Moreover, it allows for the more interesting and discontinuous demand structures, e.g. in Case 4 with RDP, which are not feasible when market is covered. We are working on this as an extension. Second, online platforms often enjoy demand-side externalities, which we ignore in our basic model. While our basic model without externalities provides a lower bound for the utility of the Incumbent, including externalities provides more interesting and discontinuous demand structures which may drive more complex pricing strategies by the platforms. We are working on this as an extension. Third, we consider myopic incumbent platform. It would be interesting to see how the platform pricing and customer data leverage change if platforms are strategic.

Fourth, we consider the case where platforms cannot price-discriminate between customers with and without data, and data porting to be free. Costly porting, whether administered by the platforms or a data broker, may provide interesting insights. In our model where prices are the same for all customers of a given platform, some customers may end up paying less or more due to the competition between platforms over a different type of customers. If price discrimination is possible, platforms may opt for more nuanced strategies to compete.

Finally, other than pricing decisions, platforms often decide on how to leverage the customer data to provide them with better service, which may be even more important than pricing decisions, as many such platforms are free to use. Platforms decide on how much effort they want to put into using and leveraging customer data. Including data leverage coefficients as decision variables in analysis would provide deeper insights into how platforms can create strategies around technologies such as customer analytics, recommendation systems, and automated agents. Moreover, it would be interesting to study the gathered data and the requirements of platforms on personal information disclosure, and how that impacts the profit of the platforms, for example in the case of targeted advertising revenue. Another interesting aspect to this problem is where platforms do not require each customer's data to create customized services, but rather they can use aggregated data from other customers for this purpose (externalities in the data). Such analyses have important implications for privacy. In our model, we abstract the customer privacy concerns and include it in their optimal level of data disclosure. It would be interesting to see what the effect of privacy would be if it is considered explicitly.

References

Acquisti A, Varian HR (2005) Conditioning prices on purchase history. *Marketing Science* 24(3):367–381.

Bélanger F, Crossler RE (2011) Privacy in the digital age: a review of information privacy research in information systems. *MIS quarterly* 35(4):1017–1042.

Bojars U, Decker S, Breslin J (2008) Social networks and data portability using semantic web technologies .

Caillaud B, De Nijs R (2014) Strategic loyalty reward in dynamic price discrimination. *Marketing Science* 33(5):725–742.

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

Christensen L, Colciago A, Etro F, Rafert G (2013) The impact of the data protection regulation in the eu. *Intertic Policy Paper, Intertic* .

Conger K (2018) How to download your data with all the fancy new GDPR tools. *Gizmodo* URL `https://gizmodo.com/how-to-download-your-data-with-all-the-fancy-new-gdpr-t-1826334079`.

De Hert P, Papakonstantinou V, Malgieri G, Beslay L, Sanchez I (2018) The right to data portability in the gdpr: Towards user-centric interoperability of digital services. *Computer Law & Security Review* 34(2):193–203.

Dewan R, Jing B, Seidmann A (2003) Product customization and price competition on the internet. *Management science* 49(8):1055–1070.

Engels B (2016) Data portability among online platforms. *Internet Policy Review* 5(2).

EPSC (2017) Enter the data economy - EU policies for a thriving data ecosystem. *European Political Strategy Centre Strategic Notes* 21.

Farrell J, Saloner G (1985) Standardization, compatibility, and innovation. *the RAND Journal of Economics* 70–83.

Farrell J, Saloner G (1986) Installed base and compatibility: Innovation, product preannouncements, and predation. *The American economic review* 940–955.

Fialová E (2014) Data portability and informational self-determination. *Masaryk UJL & Tech.* 8:45.

Gal-Or E, Gal-Or R, Penmetsa N (2018) The role of user privacy concerns in shaping competition among platforms. *Information Systems Research* .

Gehrig T, Shy O, Stenbacka R (2011) History-based price discrimination and entry in markets with switching costs: a welfare analysis. *European Economic Review* 55(5):732–739.

Gehrig T, Stenbacka R (2007) Information sharing and lending market competition with switching costs and poaching. *European Economic Review* 51(1):77–99.

Hotelling H (1929) Stability in competition. *The Economic Journal* 39(153):41–57.

Ip C (2018) Who controls your data? *Engadget* URL `https://www.engadget.com/2018/09/04/who-controls-your-data/`.

Kim SJ (2017) Privacy, information acquisition, and market competition .

Koh B, Raghunathan S, Nault BR (2017) Is voluntary profiling welfare enhancing? *Management Information Systems Quarterly* 41(1):23–41.

Lam WMW, Liu X (2018) Does data portability facilitate entry? *45th Conference of the European Association for Research in Industrial Economics.*

Nault BR, Vandenbosch MB (2000) Disruptive technologies—explaining entry in next generation information technology markets. *Information Systems Research* 11(3):304–319.

Razmerita L, Jusevičius M, Firantas R (2009) New generation of social networks based on semantic web technologies: the importance of social data portability. *Proceedings of the International Workshop on Adaptation and Personalization for Web*, volume 2, 79–87.

Salop SC (1979) Monopolistic competition with outside goods. *The Bell Journal of Economics* 141–156.

Shirazi MN, Kuan HC, Dolatabadi H (2012) Design patterns to enable data portability between clouds' databases. *Computational Science and Its Applications (ICCSA), 2012 12th International Conference on*, 117–120 (IEEE).

Smith HJ, Dinev T, Xu H (2011) Information privacy research: an interdisciplinary review. *MIS quarterly* 35(4):989–1016.

Swire P, Lagos Y (2012) Why the right to data portability likely reduces consumer welfare: antitrust and privacy critique. *Md. L. Rev.* 72:335.

Taylor CR (2004) Consumer privacy and the market for customer information. *RAND Journal of Economics* 631–650.

Zhu F, Iansiti M (2012) Entry into platform-based markets. *Strategic Management Journal* 33(1):88–106.