Standing on the Shoulders of Web Giants: The Economic Effects of Personal Data Markets∗

Bastian Haberer†, Jan Krämer‡, Daniel Schuur§

First version: August 2018
This version: April 18, 2019

Abstract

Internet users create a wealth of (personal) data when using content and service providers (CSPs), who then sell this data on the data market. Personal data markets (PDMs) encourage users to port their data from CSPs and enable them to sell their data directly. In Europe, this is further facilitated by the right to data portability introduced by the General Data Protection Regulation. Based on a game-theoretic model, we investigate how a PDM affects the CSP’s incentives to invest in the quality of its service, and its pricing strategy for consumers. We identify three, partially countervailing, economic effects that govern the relevant strategic trade-offs in an Internet market with PDM. On the one hand, the presence of a PDM will induce the CSP to reduce the quality of its service, (1) because it faces lower revenues on the data market (competition effect), and (2) because it can free-ride on the PDM-induced data creation incentive for users (displacement effect). On the other hand, a PDM will lead to an increase in the CSP’s quality, (3) because the CSP can partially appropriate the additional consumer surplus that has been created by the PDM through an increase in its price for the service (appropriation effect). The strengths of these effects depend crucially on the PDM’s efficiency on the data market, and consumers will only be able to benefit if the PDM can sell data more efficiently than the CSP.

Keywords: data markets, data sharing, data portability, intellectual property, user empowerment, digital economy

JEL classification: L13, L15, L86

∗Bastian Haberer and Daniel Schuur acknowledge funding from the Bavarian State Ministry of Science and the Arts in the framework of the Centre Digitisation.Bavaria.
†Research Group Data Policies, University of Passau, e-mail: bastian.haberer@uni-passau.de.
‡Chair of Internet and Telecommunications Business, University of Passau, e-mail: jan.kraemer@uni-passau.de.
§Research Group Data Policies, University of Passau, e-mail: daniel.schuur@uni-passau.de.
1 Introduction

The idea of personal data markets (PDMs) dates back to Laudon (1996), who envisioned the creation of a national information market, where data subjects can deposit their information in bank-like institutions and are compensated for the use of their data. However, only recently have such PDMs emerged in practice. Start-ups like Datacoup or people.io enable users to collect, store, aggregate and commercially exploit their personal (usage) data which they have already revealed when using the online platforms of web giants like Facebook or Google. In this vein, PDMs aim at empowering users by enabling them to transfer and trade their own personal data created at content and service providers (CSPs). For instance, at people.io users are remunerated in exchange for granting access to their e-mail account and at Datacoup consumers can aggregate, connect and store personal data from various services in order to be able to sell these data to data brokers (Datacoup, nd). Consumers may, for example, store personal financial data (e.g., credit and debit card accounts) together with their social media content (e.g., Facebook or Twitter) and then sell access to the entire data set. PDMs are therefore considered by many policy makers and Internet advocates as the silver bullet of the data economy, claiming that they balance the trade-off between facilitating third-parties’ access to valuable personal information and at the same time providing control and sufficient compensation to data subjects, i.e., the users (European Data Protection Supervisor, 2016).

The PDM business model relies on consumers’ ability to extract their personal data from online CSPs and use this data as an input for the PDM. The new right to data portability in the European Union, which is established in Article 20 of the General Data Protection Regulation that came into effect in May 2018, guarantees for the first time that consumers are indeed able to transfer data between online services in a machine-readable format at relatively low transaction cost (European Commission, 2016).\textsuperscript{1} In reverse, PDMs could likely foster the incentives for users to actually exercise their right to data portability by offering users an explicit monetary reward for doing so.

\textsuperscript{1}Whereas it may have been technically possible to access usage data from CSPs through Application Programming Interfaces (APIs) or by providing the PDM with the login credentials to the CSP’s service, in the past transaction costs and legal uncertainties (Koebler, 2016) constituted significant obstacles to widespread adoption.
Whereas legal aspects of the GDPR and the consequences on data-driven business models are extensively discussed in academia and practice, economic analyses on the impact of a right to data portability are scarce. In particular, the interplay between data portability and personal data markets has so far not been explored. In this study, we investigate the impact of PDMs on the Internet ecosystem on the basis of a game-theoretic model. More specifically, we consider a market setting, where a CSP offers consumers a service and invests in the quality of this service. The CSP may also charge a price or pay a subsidy to consumers for using the service. A higher consumption of the service allows the CSP to collect more user data, which it then can sell to data brokers or advertisers. In this market setting, we fully characterize equilibrium outcomes with and without a PDM. In particular, we analyze the effects of a PDM on a CSP’s incentive to invest in quality and to raise its price, as well as its impact on firms’ profit and consumer surplus. Most notably, we identify three main economic effects: First, a PDM may diminish a CSP’s revenues from selling data as it free rides on the CSP’s quality investment and introduces competition on the data market. This lowers the CSP’s incentives to invest in the quality of its service. Second, remuneration of consumers may encourage them to create more data, everything else being equal. This can be strategically exploited by the CSP, as it allows for the displacement of costly quality investments by monetary rewards paid out by the PDM in order to stimulate data creation. Third, remuneration raises consumers’ surplus. In consequence, the CSP is able to increase its own price for consumers if the PDM is willing to pay high rewards to consumers, thus leading to a (partial) appropriation of consumers’ rewards. This, in turn, incentivizes the CSP to raise its quality investment, countervailing the two other effects. Whereas the competitive effect is well-known, the displacement and appropriation effect have so far not been identified in the academic literature, but bear important managerial and policy implications. In conclusion, we find that a CSP may be able to increase its profit with a PDM, but only if the PDM is sufficiently efficient on the data market, such that it is willing to pay high rewards to consumers. Moreover, consumers can only benefit from a PDM if it has an actual competitive advantage over the CSP on the data market, i.e., the PDM is able to generate higher revenues for the same input data set. We show that these results are robust for market settings with
a competing CSP, in the presence of transaction costs for consumers to adopt the PDM and for endogenous investments in data analytics technology.

The remainder of this paper is structured as follows: First, we review the related literature. We then present our game-theoretic model and derive the equilibrium outcomes for the scenarios with and without a PDM. Afterwards, we evaluate the impact of a PDM by investigating firms’ equilibrium strategies, market outcomes and consumer welfare. We then show that our main results hold across various extended model scenarios. Finally, we conclude by discussing managerial and policy implications as well as limitations of our model.

2 Related Literature

Our study on PDMs is related to the literature on trading mechanisms for personal data as well as the empirical literature on the value of personal information. With respect to the latter, Godel et al. (2012) provide an overview of studies that have attempted to measure consumers’ willingness to pay for privacy, or in reverse, the minimum price for which consumers are willing to sell their personal information. These studies show that a majority of consumers is willing to sell personal data to third parties or disclose data publicly in return for a monetary reward. Based on this insight, studies have further investigated specific trading mechanisms that allow individuals to sell their data. For example, Rice et al. (2004) conceive an intermediary which is able to distort data and compensates data subjects according to distortion levels. Aperjis and Huberman (2012) find that paying individuals according to their risk attitude might give companies access to cheap and unbiased data sets. In a setting where user information has positive externalities on others, Chessa and Loiseau (2015) suggest to rely on an intermediary to find an appropriate level of data disclosure and user compensation. Thereby users may benefit from the positive externalities of their disclosed data. In our study, we further investigate the conditions when consumers benefit from selling their data, if the PDM is paying consumers a reward that is proportional to the amount of data and if at the same time consumers value service quality of CSPs.

Furthermore, our study contributes to the debate on the economic consequences of the Eu-
ropean General Data Protection Regulation (GDPR) and especially the Right to Portability, established in Article 20 of the Regulation. PDMs are closely related to Personal Information Management Systems (PIMS), which the European Commission (2017) has referred to as possible technical and economic instruments to implement the right to data portability. In general, PIMS are understood as a means to facilitate the control over data (see, e.g., European Data Protection Supervisor, 2016). From a technical perspective, Abiteboul et al. (2015) characterize PIMS as user-controlled servers which run specific services selected by the user and store and process the user’s data. Thus, PIMS can be used to link and enrich data, that consumers have created at CSPs as a by-product of their usage. In this vein, PIMS may be used as a technical infrastructure to facilitate the trading of personal data at external market places or may even be utilized as PDMs, where trades between consumers and firms can be executed without an additional intermediary. Outside of the European Union, initiatives like Mydata (Honey et al., 2016) in the United States have aimed to facilitate data provision. Moreover, the data portability concept itself has also been discussed in the United States (Macgillivray and Shambaugh, 2016). However, economic analyses of the GDPR and the right to data portability are rare. Wohlfarth (2019) shows that data portability is likely to have a positive effect on market entry by new firms, but consumers may suffer due to more data collection by the incumbent in anticipation of the entry. Assuming the same market setting, Lam and Liu (2018) show that data analytics technology can have a significant impact on consumers’ ability to switch and transfer data between the incumbent and the entrant and thus on the economic outcome for consumers. However, neither of these studies considers business models that remunerate consumers for the transfer of their data, which may, though, significantly affect consumers’ behavior. In fact, we show that in a market setting, where consumers value service quality, a positive reward for consumers is a necessary condition to make them better off with data portability. We show that this is also the case, if firms are able to invest in data analytics technology.

Finally, as personal data is often created as a byproduct of service usage (see Lefouili and Toh, 2018), a PDM may not only affect a user’s decision to provide data, but may also have significant effects on CSPs that provide these services and rely on the monetization of
such data (e.g., through sales to data brokers or targeted advertising). Therefore, our study is related to the literature on the advantages and potential problems of treating privacy as (intellectual) property (see, e.g., Samuelson, 2000; Schwartz, 2004). In this context, Duch-Brown et al. (2017) analyze the economic consequences of distributing rights between the creators of databases and the data subjects. More specifically, Zhu et al. (2008) investigate the incentives of a dataset’s creator to bear the fixed costs that stem from the creation of the database if another firm can reuse the contained data. They find that a low level of differentiation between the original dataset and the dataset of a free-riding secondary user might discourage the former to create the dataset in the first place. Compensation of the creator for the reuse then may in some cases help to overcome this problem. As noted by Swire and Lagos (2013) the right to data portability may exert a similar negative effect on firms’ innovation incentives and their willingness to compile data sets, because data portability may threaten a firm’s business model to monetize its unique data set. We show that this free riding in fact occurs in the context of data portability and a PDM. However, this competition effect may be reinforced or countervailed by two additional effects if consumers are remunerated through an explicit reward mechanism.

3 Monopolistic CSP without Personal Data Market

3.1 The Model

In order to evaluate the effect of a PDM on market outcomes, we first consider a benchmark scenario with a monopolistic CSP, but without PDM. We denote this scenario by superscript $M$. The basic set-up in this scenario is as follows. The CSP has two streams of revenue. First, it can charge a price for the usage of its service (which can also be zero or negative) to consumers. Second, it can make an additional revenue by selling the consumers’ data to data brokers. The CSP can strategically set both, the quality of its service, as well as the usage price in order to maximize total profits. The higher the CSP’s quality level, or the lower its price, the more consumers will use the service, and hence, more data can be sold. This basic setup is visualized in the left-hand panel of Figure 1. The detailed model set-up
is presented below.

**Without Personal Data Market (M)**

<table>
<thead>
<tr>
<th>Content &amp; Service Provider</th>
<th>Data Brokers</th>
<th>Consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Revenue</td>
<td></td>
</tr>
</tbody>
</table>

**With Personal Data Market (MP)**

<table>
<thead>
<tr>
<th>Content &amp; Service Provider</th>
<th>Data Brokers</th>
<th>Consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Revenue</td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>Revenue</td>
<td>PDM</td>
</tr>
</tbody>
</table>

Figure 1: Market structure in the base model without (left panel) and with (right panel) PDM.

**Consumers:** We consider a unit mass of homogeneous consumers. In the benchmark scenario, consumers have utility

\[
U^M(x, p, q) = \begin{cases} 
q x - \frac{x^2}{2} - px & \text{when using the CSP,} \\
0 & \text{otherwise,}
\end{cases}
\]

where \(x\) is the usage level of the CSP’s service, \(q \geq 0\) is the quality offered by the CSP and \(p\) is the price that the CSP demands for using its service. Note that we allow for cases where the CSP demands a negative price, i.e., where it compensates consumers for using its service.

Consequently, a consumer will use the CSP’s service only (i.e., choose \(x > 0\)) if her utility is positive. In this context, it is useful to think of \(x\) as the time spent on a platform. We make the assumption that utility increases in \(x\), but that marginal utility decreases in \(x\). Therefore, there exists an optimal usage level \(x\), depending on \(q\) and \(p\) that users spent at the CSP (see Casadesus-Masanell and Hervas-Drane (2015), and Lefouili and Toh (2018) for
a similar approach). Furthermore, we assume that the amount of data that can be collected by the CSP is proportional to the consumer’s usage level. For simplicity, we measure the amount of data provided to a platform directly by $x$. This modeling approach captures the characteristics of many data-driven platform services, where usage and data provision are inevitably intertwined. Note that, since consumers are homogeneous, all have the same privacy-related costs when using the CSP, which we normalize to zero.

**CSP:** There is a monopolistic CSP which makes revenues from two different sources. On the one hand, it can charge consumers of its service a price $p > 0$. On the other hand, it can sell the data generated by consumers when using its service to *data brokers*. Data brokers, in turn, monetize the data by offering personalized advertisements or advanced data analytics services to its customers. We do not model the market for data explicitly, but assume that a monopolistic CSP can sell its user data $x$ for $d(x) = dx$.

The CSP has two instruments to control the amount of data that users provide. First, the CSP can invest in the quality $q$ of this service. We assume the investment costs to be $C(q) = q^{1/2}$, such that they are increasing and convex in $q$. A higher quality increases consumers’ valuation for the service and, thus, increases the time spent at the CSP’s website or application. Second, the CSP can influence usage through the price $p(x) = px$. Everything else equal, a lower price will increase usage. Note that the CSP may even choose not to charge consumers, but to subsidize usage by setting a negative price. In practice, this does not have to be a direct monetary transfer, but could also be achieved, for example, through coupons.

Taken together, the CSP makes a profit of

$$\Pi^M(x, p, q) = d(x) + p(x) - C(q) = x(d + p) - \frac{q^2}{2}.$$  

**Timing:** In the benchmark scenario, the timing is as follows:

*Stage 1:* The CSP invests in quality, $q$.

*Stage 2:* The CSPs sets the price, $p$.

*Stage 3:* Consumers choose their usage level, $x$. 

8
3.2 Equilibrium Derivation

We proceed by backward induction in order to determine the subgame-perfect Nash equilibrium. Knowing the quality and price of the content provider, in Stage 3 consumers choose their optimal level of usage $x$ and, thereby, determine the amount of data they create for the CSP:

$$x^M(p, q) = \arg \max_{x \in \mathbb{R}_+} [U^M(p, q)] = q - p.$$

In Stage 2, investment costs with respect to quality are already sunk. Taking the effects on consumers’ usage decision into account, the CSP chooses the price that maximizes its Stage 2 profit, given $q$:

$$p^M(q) = \arg \max_{p \in \mathbb{R}_0} [\Pi(x^M, p, q)] = \frac{q - d}{2}.$$

In Stage 1, the CSP determines the quality level of its service. When maximizing its profit over $q$, it anticipates the effect of its choice on its own optimal price level $p^M(q)$ in Stage 2, as well as on the optimal usage level of consumers in Stage 3 $x^M(q)$:

$$q^M = \arg \max_{q \in \mathbb{R}_0} [\Pi^M(x^M, p^M, q)] = d.$$

Consequently, in equilibrium, we obtain:

$$q^M = d, \quad p^M = 0, \quad x^M = d, \quad \Pi^M = \frac{d^2}{2}.$$

4 Monopolistic CSP with Personal Data Market

4.1 The Model

We now extend the basic model introduced above in order to study how a PDM would change market outcomes. The PDM can also sell the CSP consumers’ data (the data generated while using the CSP) to data brokers. That is, it competes with the CSP on the data market. In order to receive access to consumers’ data, the PDM chooses a positive reward that it is willing to pay in return for their data. A higher reward incentivizes consumers to use the
CSP more and thereby to “produce” more data. We denote this scenario by superscript $MP$ and it is visualized in the right hand panel of Figure 1. The model with PDM is described in detail below.

**PDM:** For now, we assume that participation at the PDM does not incur any fixed or transaction costs, such that all consumers use the PDM, provided it offers a positive reward for data. In particular, the PDM offers the consumers a reward of $r(x) = rx$ for the provided data. In return all consumers port their usage data $x$ from the CSP to the PDM, e.g., by uploading their ported data, providing the PDM with their login credentials, or through the use of an application programming interface (API). Thus, the PDM has access to the same data $x$ as the CSP. Taken together, the profit function of the PDM is

$$\Gamma^{MP}(x, r) = \delta \varrho d(x) - r(x) = (\delta \varrho d - r) x.$$ 

Thereby, $\delta \in (0, 1]$ models the intensity of competition on the data market. Due to the advent of the PDM, the CSP cannot sell the user data to the data brokers exclusively anymore, which will likely reduce revenues. Moreover, we assume that the PDM makes $\varrho \in [0, \infty)$ times the profit from selling data than the CSP. This allows us to capture the efficiency of the PDM on the data market. Note that for $\varrho > 1$ we are able to model the situation in which the PDM is more efficient than the CSP and can make more revenues on the data market (e.g., because it has complementary user data from other sources that make the data more valuable), as well as for $\varrho < 1$, we can model situations in which the PDM is less efficient and makes less revenues on the data market (e.g., because it lacks the data analytics capabilities to refine the raw data received).

**CSP:** The profit function of the CSP is the same as without PDM, with the exception that the revenues from selling data are reduced by a factor of $\delta$.

$$\Pi^{MP}(x, p, q) = \delta d(x) + p(x) - C(q) = x (\delta d + p) - \frac{q^2}{2}.$$
**Consumers:** The utility function of consumers is the same as without PDM, but now consumers can additionally receive a reward of \( r(x) = rx \) from the PDM for providing it with data:

\[
U^{MP}(x,p,q,r) = \begin{cases} 
q \frac{x^2}{2} - px + rx & \text{when using the CSP and PDM}, \\
q \frac{x^2}{2} - px & \text{when using only the CSP}, \\
0 & \text{otherwise},
\end{cases}
\]

Consequently, we assume that for \( r > 0 \) all consumers use the PDM.

**Timing:** With the PDM, the timing is as follows:

1. **Stage 1:** The CSP invests in quality, \( q \).
2. **Stage 2:** The CSPs sets the price, \( p \), and the PDM the reward, \( r \).
3. **Stage 3:** Consumers choose whether to use the PDM and choose their usage level, \( x \).

### 4.2 Equilibrium Derivation

We again use backward induction to derive the subgame-perfect Nash equilibrium. In **Stage 3** consumers determine their usage level by maximizing utility given \( q, p \) and \( r \). Clearly, if \( r \leq 0 \) no consumer would use the PDM and the same results as in the benchmark scenario would arise. Assuming \( r > 0 \), consumers maximize utility by choosing

\[
x^{MP}(p,q,r) = \arg \max_{x \in \mathbb{R}_+^*} [U^{MP}(x,p,q,r)] = q - p + r.
\]

In **Stage 2** the CSP as well as the PDM anticipate consumers’ usage decisions and simultaneously set \( p \) and \( r \) to maximize their respective Stage 2 profits. From the systems of equations given by the first-order conditions \( \frac{\partial \Pi^{MP}(x^{MP},p,q)}{\partial p} = 0 \) and \( \frac{\partial \Gamma^{MP}(x^{MP},r)}{\partial r} = 0 \), we can derive the following equilibrium candidate in **Stage 2**, given the CSP’s quality investment in
Stage 1:

\[ r^{MP}(q) = \arg\max_{r \in \mathbb{R}_0^+} \left[ \Gamma(x^{MP}, r(p^{MP})) \right] = \begin{cases} -\frac{q + \delta d (2\varrho - 1)}{3} & \text{if } q < \tilde{q} := \delta d (2\varrho - 1) - 3\epsilon, \\ \epsilon & \text{if } q \geq \tilde{q}, \end{cases} \]

\[ p^{MP}(q) = \arg\max_{p \in \mathbb{R}_0^+} \left[ \Pi(x^{MP}, p(r^{MP}), q) \right] = \begin{cases} \frac{q - \delta d (\varrho - 2)}{3} & \text{if } q < \tilde{q}, \\ \frac{q - \delta d + \epsilon}{2} & \text{if } q \geq \tilde{q}. \end{cases} \]

In general, the PDM’s optimal reward \( r^{MP} \) is decreasing in the quality investment \( q \) by the CSP in Stage 1. For \( q > \tilde{q} \), the PDM would even want to set a negative reward, \( r < 0 \), according to this equilibrium candidate. However, this cannot constitute an equilibrium, since then, consumers would choose not to visit the PDM and the PDM would make zero profit. Thus, in this case, another equilibrium candidate emerges in which the PDM sets the smallest possible reward, \( r^{MP} = \epsilon \), to ensure consumers’ participation.

In Stage 1, the CSP anticipates the ensuing prices and usage levels and chooses how much to invest in quality. The optimal quality level is determined by maximizing profits under consideration that the ensuing price structure in Stage 2 depends on whether \( q < \tilde{q} \).

Specifically, the CSP may adjust its quality level to accommodate to the existence of the PDM. This candidate equilibrium investment is readily given by the first-order condition \( \frac{\partial \Pi^{MP}}{\partial q} = 0 \) which yields \( q^{\text{MPE}} = \frac{2}{7} \delta d (1 + \varrho) \). Or the CSP can choose a monopoly-like quality level of \( q^{\text{MP}} = \delta d + \epsilon \), which induces the PDM to set a minimum reward \( \epsilon \). It can then be shown that the profit-maximizing quality choice is given by

\[ q^{MP} = \arg\max_{q \in \mathbb{R}_0^+} \left[ \Pi^{MP} \left( x^{MP}, p^{MP}, q \right) \right] = \begin{cases} \frac{2}{7} \delta d (1 + \varrho) & \text{if } \varrho > \tilde{\varrho} := \sqrt{\frac{2}{7} \sqrt{\frac{(\delta d + \epsilon)^2}{\varrho^2}} - 1}, \\ \delta d + \epsilon & \text{if } \varrho \leq \tilde{\varrho}. \end{cases} \]

For \( \varrho < \tilde{\varrho} \), it is profitable for the CSP to set \( q^{\text{MP}} \) in order to marginalize the PDM in Stage 2. In contrast, for \( \varrho \geq \tilde{\varrho} \), the CSP chooses a quality level \( q^{\text{MP}} \), which allows the PDM to set a positive reward \( r > \epsilon \) according to the PDM’s unconstrained best response function in Stage 2.

Therefore, there are two types of subgame-perfect equilibria in pure strategies that may
arise depending on the efficiency of the PDM, $\varrho$:

- **POSITIVE-INCOME EQUILIBRIUM (for $\varrho > \overline{\varrho}$):**

  The positive-income equilibrium is characterized by the fact that the total rewards offered by the PDM exceed the cost of participation at the PDM.

  \[
  q^M_p = \frac{2}{7}\delta d (1 + \varrho), \quad p^M_p = \frac{1}{7}\delta d (3\varrho - 4), \quad r^M_p = \frac{1}{7}\delta d (4\varrho - 3), \quad x^M_p = \frac{3}{7}\delta d (1 + \varrho),
  \]

  \[
  \Pi^M_p = \frac{1}{7}\delta^2 d^2 (1 + \varrho)^2, \quad \Gamma^M_p = \frac{9}{49}(\delta d + \delta \varrho d)^2.
  \]

- **MINIMUM-INCOME EQUILIBRIUM (for $\varrho \leq \overline{\varrho}$):**

  The minimum-income equilibrium is characterized by the fact that the total rewards offered by the PDM just offset the cost of participation at the PDM.

  \[
  q^M_{\epsilon} = \delta d + \epsilon, \quad p^M_{\epsilon} = \epsilon, \quad r^M_{\epsilon} = \epsilon, \quad x^M_{\epsilon} = \delta d + \epsilon,
  \]

  \[
  \Pi^M_{\epsilon} = \frac{1}{2}(\delta d + \epsilon)^2, \quad \Gamma^M_{\epsilon} = (\delta d + \epsilon)(\delta \varrho d - \epsilon).
  \]

  Note that $\overline{\varrho} \rightarrow \sqrt{\frac{7}{2}} - 1 \approx 0.871$ for $\epsilon \rightarrow 0$. Thus, the minimum-income equilibrium only emerges if the PDM is not sufficiently efficient in selling user data vis-à-vis the CSP. In contrast, if the PDM is sufficiently efficient ($\varrho > 0.871$, i.e., even though it may still be less efficient than the CSP), the CSP will always accommodate its quality level to the advent of the PDM, and the PDM will always offer users a positive reward in equilibrium.

5 **Comparison of market outcomes with and without Personal Data Market**

Based on the insights from the previous two sections, we can now compare how a PDM affects the market outcome. To this end, we investigate the impact of the PDM on the CSP’s quality, which allows us to determine three fundamental economic trade-offs that govern the strategic interplay between the CSP and the PDM.
5.1 CSP’s incentives to invest in quality

**Competition Effect:** In both, the minimum-income equilibrium and the positive-income equilibrium, consumers will port their user data to the PDM. We have assumed that this leads to a decrease in the CSP’s marginal revenues from selling data by a factor of $\delta \leq 1$, because the CSP now competes with the PDM, which is in possession of the same data set and may sell this data to the same data brokers. This reduction in marginal revenue from selling data has an immediate impact on the CSP’s marginal incentive to stimulate data creation through its service quality. Indeed, one can show that with increasing competitive pressure from the PDM, the CSP reduces its investments in quality, i.e., $\frac{\partial q}{\partial \delta} < 0$.

**Insight 1 (Competition Effect)** *In the presence of a PDM, the CSP reduces its investments in quality due to the increased competition on the data market, everything else being equal. The stronger the competition, the stronger the reduction in quality investments ($\frac{\partial q}{\partial \delta} < 0$).*

The competition effect is most evident in the minimum-income equilibrium, where the PDM has no other significant effect on the CSP, because it offers only a minimum reward for data to consumers. That is, in the minimum-income equilibrium the PDM’s reward does not significantly influence consumers’ usage of the CSP, everything else being equal. Recall that in this case the CSP sets a monopoly-like quality level of $q = \delta d + \epsilon$. For $\delta = 1$ (and $\epsilon \rightarrow 0$), i.e., without any competitive impact of the PDM, the CSP would set the same quality level as without the PDM; but for any $\delta < 1$ it sets a lower quality level. This is demonstrated in Figure 2a for $q < \overline{q}$.

The competition effect is also present in the positive-income equilibrium (i.e., for $q \geq \overline{q}$), but here two additional effects arise, so that the overall quality level can be higher or lower than in the minimum-income equilibrium. We describe these effects in turn.

**Displacement Effect:** In the positive-income equilibrium, i.e., for $q > \overline{q}$, the PDM is willing to remunerate consumers for transferring their data from the CSP to the PDM by offering a reward that significantly exceeds $\epsilon$. It does so, to incentivize users to create more data (as $\frac{\partial \epsilon}{\partial r} > 0$) than they would otherwise do, given the price and quality level of the CSP. However,
notice that the PDM’s reward, $r$, and the quality, $q$, of the CSP are strategic substitutes, as $\frac{\partial r}{\partial q} < 0$. Thus, in the positive-income equilibrium the CSP anticipates that the negative effect of a lower quality level on usage will be partially compensated by the PDM, because it will in turn offer a higher reward to increase usage again. In other words, knowing that the PDM will set a positive reward (in Stage 2), induces the CSP to set a lower quality (in Stage 1) than it would in the minimum-income equilibrium, everything else equal.

**Insight 2 (Displacement Effect)** In the positive-income equilibrium (i.e., for $\varrho > \overline{\varrho}$), the CSP has a lower incentive to invest in quality, because the PDM offers consumers a positive reward for data, everything else being equal. This occurs even if the PDM does not compete with the CSP on the data market ($\delta = 1$).

To see this formally, assume the CSP’s price is fixed at $p_0$, but that $q$ and $r$ are chosen endogenously according to the best-response functions. Now let us, hypothetically, compare the qualities that would emerge in the positive-income equilibrium and the minimum-income equilibrium at $p_0 = 0$. It follows that $\Delta q = q^\text{MP}(p = 0) - q^\text{MP}(p = 0) = -\frac{d\delta}{2} - \epsilon < 0$. That is, absent additional price effects, the CSP will lower its quality below the level set in the minimum-income equilibrium. Also notice that the displacement effect is distinct from the competition effect and occurs even when the competition effect is absent, i.e., $\delta = 1$.

Therefore, the displacement effect can be best observed at the threshold $\overline{\varrho}$, when the market tips from the minimum-income equilibrium to the positive-income equilibrium (see Figure 2a). At this threshold, the CSP’s price is close to zero, and the quality drop by about $\frac{d\delta}{2}$.

**Appropriation Effect:** In the positive-income equilibrium (i.e., for $\varrho > \overline{\varrho}$), the reward does not only depend on the CSP’s service quality, $q$, but also on the efficiency of the PDM on the data market, $\varrho$. When the PDM becomes more efficient and can reap higher revenues from selling data, it increases consumers’ reward for data in the positive-income equilibrium, i.e.,

---

2In fact, the CSP’s price is slightly negative at this threshold and the quality drop is therefore slightly higher than $\frac{d\delta}{2}$. This is because the CSP finds it more profitable to reduce the quality level even further (in Stage 1) than it would at a price of $p = 0$, in order to induce the PDM to set an even higher reward, i.e., to maximize the displacement effect. To counterbalance this decrease in quality and its negative effect on usage, the CSP subsidizes the usage in Stage 2 through a negative price.
∂r > 0. At the same time, the PDM’s reward and the CSP’s price are strategic complements, i.e., \( \frac{\partial p}{\partial r} > 0 \) and \( \frac{\partial p^{MP}}{\partial \varrho} > 0 \) for \( \varrho > \overline{\varrho} \). Hence, when the PDM offers a larger reward for consumers, the CSP will optimally increase its price in order to appropriate part of the additional consumer surplus (see Figure 2b). Realizing that such appropriation of the PDM’s efficiency gain is possible (in Stage 2), gives the CSP an incentive to increase its quality (in Stage 1), everything else equal. In fact, notice from (1) that \( \frac{\partial p^{MP}}{\partial \varrho} > 0 \) in the positive-income equilibrium (see also Figure 2a). This appropriation effect stems from the fact that the CSP does not only anticipate how its investment will influence the PDM’s reward, but also how it will affect its own price setting. Therefore, the appropriation effect can only exist if the CSP is able to set a price on the consumer market.

**Insight 3 (Appropriation Effect)** In the positive-income equilibrium (i.e., for \( \varrho > \overline{\varrho} \)), the CSP has an incentive to increase its quality investment as the PDM becomes more efficient on the data market (\( \varrho \) increases), because it is able to appropriate some of the additional consumer surplus created by the PDM through a higher price \( p \).

![Figure 2: Quality, prices and rewards (for \( \delta = 0.7, d = 1, \epsilon = 0 \)].](image)

5.2 Impact of the PDM on market outcomes

**Quality:** As shown in the previous subsection, there are three countervailing effects of the PDM on the CSP’s quality choice. On the one hand, the competition effect and the
displacement effect lower the CSP’s investment in quality, on the other hand, the appropriation effect has a positive impact on the quality choice of the CSP. Both, the appropriation effect and the displacement effect are only present in the positive-income equilibrium, but not in the minimum-income equilibrium (where the PDM’s reward and the CSP’s price are close to zero). Furthermore, the strength of these effects varies and crucially depends on PDM’s efficiency on the data market $\varrho$. While the displacement effect is negative and independent of $\varrho$, the appropriation effect is positive and increasing in $\varrho$. Consequently, for high levels of $\varrho > \varrho^* := \frac{5}{2}$, the appropriation effect can fully offset the displacement effect, leading to investments above the level in the minimum-income equilibrium. For even higher levels of $\varrho > \varrho^\theta := \varrho \leq \frac{7}{45} - 1$, the appropriation effect can even offset both, the competition and the displacement effect, leading to higher investments in quality than without the presence of a PDM. This is illustrated in Figure 2a for a numerical example.

**Proposition 1 (Quality Investment)** The impact of the PDM on the CSP’s investment in quality depends on the efficiency ($\varrho$) of the PDM on the data market. If the PDM is of low efficiency ($\varrho < \varrho^\ell$), the CSP’s quality is reduced by factor $\delta$ due to the competition effect. At intermediate efficiency levels ($\varrho^\ell < \varrho < \varrho^*$), the CSP’s quality is further reduced due to the dominance of the displacement effect over the appropriation effect. At high levels of efficiency ($\varrho > \varrho^\theta$), the appropriation effect dominates the competition and displacement effect, leading to a higher quality of the CSP than without the presence of a PDM.

**Prices** Without a PDM, the CSP finds it optimal not to charge consumers for its service and to reap profits via the data market instead. With a PDM, as long as the PDM is not too efficient on the data market ($\varrho < \varrho^\ell$), the minimum-income equilibrium emerges and the CSP still sets a zero price for its consumers (assuming $\epsilon \to 0$). However, in the positive-income equilibrium, i.e., when the PDM is efficient enough such that it can properly remunerate consumers for their data, the PDM departs from the zero price. In the range, where $\varrho < \varrho < \varrho^p := \frac{4}{3}$, the CSP chooses a negative price, because the appropriation effect is still relatively weak at this efficiency level, and the CSP prefers to maximize the displacement effect in order to economize on investment costs. For $\varrho > \varrho^p$ the PDM’s efficiency is so high that the
appropriation effect becomes more relevant, such that the CSP begins to charge consumers a positive price in order to appropriate part of the surplus that the PDM creates. Hence, as consumers receive a higher remuneration from the PDM, the CSP changes its business model to also generate revenues directly from consumers, instead of relying only on revenues from the data market (see also Figure 2b).

**Proposition 2 (Price structure)** If the PDM has a large competitive advantage over the CSP on the data market \( \varrho > \varrho' \), the CSP charges consumers a higher price for the usage of its service than without a PDM.

![Graphs](image)

(a) CSP and PDM profits.  
(b) Consumer surplus.

Figure 3: Profits and consumer surplus (for \( \delta = 0.7, d = 1, \epsilon = 0 \)).

**Profits:** The PDM’s profit generally increases with its efficiency, \( \varrho \). However, as illustrated in Figure 3a, the increase is not monotonic, as the PDM’s profit falls at \( \varrho = \varrho' \) when the market tips from the minimum-income to the positive-income equilibrium. This is, because the PDM has to compensate for the displacement effect through an increase in the reward paid to consumers. However, at high efficiency levels the PDM is better off in the positive-income equilibrium.

The CSP’s profit also (weakly) increases with the PDM’s efficiency. In the minimum-income equilibrium, the CSP suffers from the competition effect and its profit loss depends on \( \delta \), but is independent of \( \varrho \). In the positive-income equilibrium, the PDM’s profit increases alongside the appropriation effect, i.e., with \( \varrho \). In consequence, at high levels of
$\varrho > \varrho^{\Pi} := \sqrt{\frac{7}{23}} - 1$, the CSP can generate higher profits in the presence of a PDM due to the appropriation effect. Furthermore, it is easy to see that in the special case where there is no competitive effect ($\delta = 1$), the CSP is always weakly better off with a PDM.

**Proposition 3 (Profits)** The CSP’s profit weakly increases in the efficiency of the PDM to sell data due to the appropriation effect. The CSP makes higher profits than without a PDM if the PDM’s efficiency on the data market is relatively large ($\varrho > \varrho^{\Pi}$).

**Welfare:** Finally, we also consider the impact of the PDM on the welfare of consumers. Without the PDM, consumer surplus is

$$CS^M = U^M(x^M, p^M, q^M) = \frac{d^2}{2},$$

and with the PDM it is

$$CS^{MP} = U^{MP}(x^{MP}, p^{MP}, q^{MP}, r^{MP}) = \begin{cases} \frac{9}{58} \varrho^2 d^2 (1 + \varrho)^2 & \text{if } \varrho > \varrho, \\ \frac{1}{2} (\delta d + \epsilon)^2 & \text{if } \varrho \leq \varrho. \end{cases}$$

Therefore, consumers are better off with a PDM if and only if $\varrho > \varrho^{CS} := \frac{7}{35} - 1$. This is illustrated by Figure 3b. In general, consumer surplus is reduced in the presence of a relatively inefficient PDM due to the competition effect and the displacement effect, which both lead to a reduction in quality by the CSP. However, a relatively efficient PDM can offer high rewards, which leads to an increase in quality due to the appropriation effect. Although part of the consumers’ surplus that is generated by the PDM’s reward is appropriated again by the CSP through a positive price, consumers can still be better off overall if the PDM’s efficiency level exceeds $\varrho > \varrho^{CS}$. Note that $\varrho^{CS} / \varrho > 0$, which means that it is less likely that consumers benefit from the PDM if the PDM leads to a higher reduction in profits on the data market for the CSP.

**Proposition 4 (Consumer surplus)** A PDM can increase consumer surplus only if it is sufficiently more efficient than the CSP on the data market (i.e., $\varrho > \varrho^{CS} := \frac{7}{35} - 1$). The
higher the level of competition induced by the PDM on the data market (i.e., the lower $\delta$), the higher needs to be the PDM’s efficiency in order to benefit consumers.

The results of this section are summarized in Figure 4, which plots market outcomes with the PDM, depending on the efficiency of the PDM ($\varrho$) and the competition intensity on the data market ($1-\delta$). Consumer welfare-decreasing outcomes are denoted by prefix $A$, welfare-increasing outcomes by $B$. A low efficiency of the PDM always leads to a plain data portability market outcome without remuneration for consumers (market outcome $A_1$). A more efficient PDM is willing to offer a positive reward, but the CSP’s profit and consumer surplus still decrease in the presence of the PDM (market outcome $A_2$). In contrast, the CSP increases its profit in market outcome $A_3$, but consumers are still worse off, despite a positive reward by the PDM. Only for higher levels of efficiency or lower levels of competition intensity, consumers are able to benefit from the PDM, although the CSP offers a lower service quality (market outcome $B_1$). Finally, in market outcome $B_2$ consumer surplus, CSP’s profit and quality all increase in the scenario with PDM.
6 Model extensions

In the following, we explore several extensions of the base model to test the robustness of our main insights (see Figure 5). First, we analyze an alternative market structure with a second CSP that also competes for consumers’ attention. This allows us to study the impact of a PDM for varying degrees of competition between CSPs. Second, we consider transaction costs of consumers, which relaxes the assumption that consumers visit the PDM for any reward $r > 0$. Finally, we take into account firms’ ability to invest in data analytics technology to increase their return on the data collected from consumers. This allows us to endogenize the parameter $\varrho$ of the PDM and evaluate market outcomes based on firms’ relative ability to implement and exploit data analytics technology.

Figure 5: Model extensions.
6.1 Competing CSPs and PDM

In the main model we considered a monopoly CSP with market power. We now relax this assumption by introducing a second, competing CSP in a duopoly framework à la Singh and Vives (1984). This framework allows us to vary the degree of substitutability ($\gamma$) between the services of the two CSPs continuously, ranging from independence (resembling the case of a monopoly) to perfect substitutability (resembling perfect competition). Moreover, the competition framework rests on a representative consumer, who determines endogenously how much the service of a given CSP will be used. In this sense, the framework also reflects that users can multihome between the CSPs. Finally, the timing is consistent with the timing in the base model: In Stage 1 CSP $A$ and CSP $B$ simultaneously invest in quality. In Stage 2, the CSPs and, if present, the PDM simultaneously decide about their prices and rewards. Finally, in Stage 3, users decide how much to use each CSP and profits are realized.

**Duopolistic CSPs without PDM (D):** In line with the model of Singh and Vives (1984), a representative consumer’s utility in the absence of a PDM is:

$$u^D = q_A x_A - \frac{1}{2} x_A^2 - p_A x_A + q_B x_B - \frac{1}{2} x_B^2 - p_B x_B - \gamma x_A x_B,$$

where $\gamma \in [0, 1)$ represents the degree of substitutability between the services with $\gamma = 0$ corresponding to independent services and $\gamma = 1$ corresponding to perfectly substitutable services. Moreover, $x_i$ reflects the usage level of CSP $i$ ($i = A, B$), $p_i$ the price charged by CSP $i$, and $q_i$ the respective quality levels. Note that for $\gamma = 0$ the competitive case is reduced to the main model’s monopoly case, with two monopolists in independent markets.

While we relegate the details to Appendix B, under competition two types of (symmetric) equilibria can emerge. First, for $\gamma \leq \gamma \approx 0.83$ both CSPs will set a positive quality level,

$$q_i^D = \frac{2(2 - \gamma) d}{(2 - \gamma)(1 - \gamma) \gamma (2 + \gamma) + 4} > 0 \quad (2)$$

in Stage 1 and subsidize usage by demanding a negative price $p_i^D < 0$. As the degree of substitutability (respectively competition) between the CSPs, $\gamma$, increases, the CSPs select
a lower service quality in equilibrium (i.e., $\partial q^D/\partial \gamma < 0$) and prices decrease (i.e., $\partial p^D/\partial \gamma < 0$). Second, for very high levels of $\gamma > \bar{\gamma}$, both CSPs choose to supply a service of minimal quality ($q^D_i = 0$). In order to induce consumers to nevertheless use the service (so that they generate user data) the CSPs choose to pay them a negative price $p^D_i < 0$, and resell the data at a price of $d > -p^D_i > 0$ on the data market.

Clearly, for $\gamma > \bar{\gamma}$, even without the presence of a PDM, competition is so intense that CSPs can merely act as arbitrageurs, “buying” data from consumers and “selling” it on the data market. In order to make a meaningful comparison to the base model, we will therefore focus the subsequent analysis on the parameter range where the first type of equilibrium exists, i.e., where CSPs compete in both, quality and price.

**Duopolistic CSPs with PDM (DP):** We now introduce a PDM in this duopoly framework. The PDM collects user-generated data from both CSPs ($A$ and $B$) and offers a reward $r \geq 0$ for each data point. Thus, the representative consumer’s utility is now given by:

$$ u^{DP} = u^D + r(x_A + x_B) $$

One can show that in the presence of the PDM two types of equilibria can emerge, akin to the minimum-income equilibrium and the positive-income equilibrium in the base model (see Appendix B). When the PDM is relatively efficient on the data market ($\varrho \geq \varrho^{DP}$), a positive-income equilibrium with $r^{DP}_+ > 0$ exists. However, when the PDM is relatively inefficient ($\varrho \leq \varrho^{DP}$), then a minimum-income equilibrium exists, in which the PDM offers the minimum reward to secure user participation, i.e., $r^{DP}_- = \epsilon \approx 0$. In each type of equilibrium, both CSPs will set their quality level as follows:

$$ q^{DP}_i = \begin{cases} 
\frac{(\gamma(2\gamma-1)-5)\delta(\varrho+1)}{(2-\gamma)(\gamma-1)(\gamma-1)-(\gamma-1)+13} & \text{if } \varrho \geq \varrho^{DP} \\
\frac{2(2-\gamma)^2\delta}{(2-\gamma)(1-\gamma)(\gamma+2)+4} & \text{if } \varrho \leq \varrho^{DP}
\end{cases} \tag{3} $$

By comparing (2) with (3) it is easy to see that in the minimum-income equilibrium, given that the PDM induces additional competition on the data market ($\delta < 1$), the CSPs choose
a lower quality level in the scenario with PDM. Like in the base model, this is because in the minimum-income equilibrium only the competition effect is present. By contrast, in the positive-income equilibrium, the CSPs’ equilibrium quality levels can be higher or lower in the presence of a PDM, depending on the level of $\varphi$. For large $\varphi > \frac{2(\gamma^2-2)((\gamma-2)\gamma((\gamma-1)\gamma-7)+13)}{((\gamma-2)(\gamma-1)\gamma(\gamma+2)+4)(\gamma(2\gamma-1)-5)^2} - 1$, the appropriation effect dominates the displacement and competition effects, such that quality levels with PDM exceed those without PDM. Consequently, as the PDM has the same qualitative effect on market outcomes with duopolistic CSPs as in the base model with a monopolistic CSPs, our insights from Propositions 1 to 4 also continue to hold qualitatively.

**Proposition 5 (Competing CSPs)** Propositions 1 to 4 continue to hold qualitatively when CSPs are competing for users’ attention.

### 6.2 Transaction costs for using the PDM

So far we have assumed that consumers are willing to use the PDM for any positive reward. Yet, in practice, it is likely that consumers experience transaction costs when using and selling their data at the PDM. These transaction costs, which we will measure by $k$, may, e.g., arise from the need to set up a user account at the PDM, enter billing credentials, and transferring and possibly enriching data from the CSP. In these cases, consumers will adopt and use the PDM only if the remuneration paid out by the PDM outweighs the transaction costs, i.e., if $rx > k$. In reverse, this offers new strategic considerations for the CSP, as the PDM may remain inactive if it is unprofitable to reward consumers in excess of their transaction costs, thus restoring a monopoly position and higher revenues on the data market.

**Monopolistic CSP with PDM and transaction costs (TP):** With transaction costs consumers’ utility is given by

\[
\begin{align*}
    u_{TP} &= \begin{cases} 
    v + qx - \frac{1}{2}x^2 - px + rx - k & \text{when using the CSP and PDM,} \\
    v + qx - \frac{1}{2}x^2 - px & \text{when using only the CSP,} \\
    0 & \text{otherwise.}
    \end{cases}
\end{align*}
\]
where \( v \geq k \) is the consumers’ base utility for using the CSP. This is to ensure that consumers can always bear the transaction costs for using the PDM, even if the CSP offers zero quality, but yet face a trade-off whether to use the PDM or not. Thus, in Stage 3 consumers decide to adopt and use the PDM if and only if \( r x^{TP} > k \). In Stage 2, it is then easy to see that, for the PDM, a minimal reward of \( r = \epsilon \) will not suffice to attract consumers. Instead a reward of \( r = k + \epsilon / x^{TP} \) is necessary to induce consumers’ participation.\(^3\) On the other hand, a PDM is never willing to set a reward that exceeds its revenue on the data market, because this would lead to negative profits. Thus, the maximum reward the PDM is willing to offer is given by \( r_{\text{max}} = d \delta q \). If the PDM rewards consumers in excess of their transaction costs, it maximizes profit by offering a reward according to its unconstrained best response \( r_{\text{TP}}^{+} \), which is identical to the positive-income equilibrium in the base model.

At the same time, the CSP may pursue different strategic options when setting its price in Stage 2, which depend on the anticipated reward of the PDM, the efficiency of the PDM, and the quality set in Stage 1. In general, three price setting strategies are at the CPS’s disposal in Stage 2, which all constitute equilibrium candidates.\(^4\)

(I) First, the CSP may choose to (actively) foreclose the market for the PDM by setting a price that induces consumers not to use the PDM in Stage 3. Two subcases must be differentiated here. (Ia) Foreclosure may already occur as a side effect of monopoly price setting (without considering the PDM), i.e., \( p_{\text{Ta}}^{TP} = p^{M} \). This occurs if the PDM is relatively inefficient and transaction costs are high. (Ib) However, in cases where \( p^{M} \) does not suffice to deter entry of the PDM, the CSP may find it profitable to actively foreclose the market for the PDM. The CSP can do so, by choosing a price \( p_{\text{Ib}}^{TP} \), which is the minimum feasible price that satisfies the condition \( k > r_{\text{max}} x^{TP} = r_{\text{max}} (q - p + r_{\text{max}}) \). In both cases, these equilibrium candidates would yield a monopoly outcome.

(II) Second, the CSP could choose to (actively) accommodate the existence of the PDM by allowing it to just enter the market. In other words, the CSP would seek a minimum-income outcome in which it sets its price such that \( k = r x^{TP} \). Here three subcases must be

\(^{3}\)For simplicity, we will assume \( \epsilon \to 0 \) in the following.

\(^{4}\)See Appendix C for a formal characterization of the PDM and CSP’s equilibrium candidates in a case where all these equilibrium candidates exist.
differenced. (IIa) In the first subcase, $p_{IIa}^{TP}$ ensures that the PDM has to pay out its entire
data revenue as reward to gain access to consumers, i.e., $k = r_{max} x^{TP}$ holds, and the PDM makes zero profits. (IIb) In the second subcase, $p_{IIb}^{TP}$ induces the PDM to set a reward, which just compensates consumers for their transaction costs, but which exceeds the reward the PDM would set according to its unconstrained best response $r^{TP}(p_{IIb}^{TP})$. Thus, it holds that $k = r x^{TP}$ with $r^{TP}(p_{IIb}^{TP}) < r < r_{max}$, such that the PDM makes positive profits. (IIc) In the third subcase, $p_{IIc}^{TC}$ is set such that the PDM’s best response is just enough to ensure that consumers use the PDM. Here, it holds that $k = r^{TP}(p_{IIc}^{TC}) x^{TP}$.

(III) Third, the CSP can set a price $p_{III}^{TC}$ that allows the PDM to choose a reward according to its best response function $r^{TP}(p_{III}^{TC})$, for which consumers derive a positive-income outcome, i.e., $k < r^{TP} x^{TP}$ holds. In particular, this means that the CSP would select $p_{III}^{TC} = p^{TP}$ as in the positive-income equilibrium in the base model.

Note that, while this represents a full characterization of all possible equilibrium candidates in Stage 2, not all candidates may be feasible for all parameter constellations. Rather, for each parameter constellation, only a subset of the candidates exists, which then represents the feasible equilibrium candidates to be considered in Stage 1.

In Stage 1, depending on the efficiency of the PDM $\varrho$ and consumers’ transaction costs $k$, the CSP chooses its quality investment $q$ in anticipation of the equilibrium candidates in Stage 2, as detailed above. Thus, in principle, six different equilibrium candidate choices of $q$ exist in Stage 1, one for each Stage 2 equilibrium candidate. However, as noted above, not all equilibrium candidates are feasible for a given parameter constellation. Among the feasible equilibrium candidates, the CSP will then select the quality level that maximizes its profits. Given the existence of up to six possible equilibrium candidates, a full characterization of the optimal choices for all possible parameter constellations is very complex and not very insightful. Instead, we illustrate the possible equilibrium outcomes for a numerical example in which we vary again the the efficiency of the PDM, $\varrho$. Figure 6a illustrates the optimal quality choice as a function of $\varrho$ in the numerical example $k = 3, \delta = 0.5, d = 3, v = 3$. Figure 6b plots the corresponding equilibrium profit of the CSP.

If the PDM is very inefficient, the CSP simply chooses its monopoly quality $q^{M} = d$, 26
because in this case the PDM will not be able to attract consumers anyway (region \( Ia \)). If the PDM becomes more efficient (region \( Ib \)), the CSP reduces its quality relative to the unconstrained monopoly level in order to actively foreclose the PDM. In this case, the CSP anticipates that a lower service quality cannot be offset by a higher reward of the PDM due to the PDM’s limited ability to generate revenues on the data market. Figure 6b illustrates that this foreclosure strategy becomes more costly as the PDM becomes more efficient, because the CSP has to artificially reduce its quality even further. The sharp increase in the CSP’s quality between region \( Ib \) and region \( IIa \) (see Figure 6a) highlights a fundamental change in the CSP’s strategy. Due to the higher efficiency of the PDM, foreclosure is so costly that, instead, it is now more profitable for the CSP to accommodate to the existence of the PDM. Thus, in region \( IIa \), the CSP sets its quality, such that the Stage 2 outcome ensures that the PDM must offer its entire revenue from the data market as reward in order to attract consumers. In consequence, this provides the highest possible stimulus for consumers to increase their usage of the CSP’s service. Within region \( IIa \), the CSP then again reduces its quality investment for a more efficient PDM, due to the PDM’s increased ability to pay a higher reward to consumers. This resembles the displacement effect. Proceeding to region \( IIb \), the quality decrease of the CSP flattens out and remains constant for increasing levels of the PDM’s efficiency. Within this region, the CSP is better off with sticking to its profit-maximizing quality from region \( IIa \) than to further reduce quality and induce the PDM to
pay out its entire data revenue to consumers. In fact, the outcome in this region resembles
the minimum-income equilibrium in the base model, where the PDM chooses its reward to
precisely compensate consumers for their transaction costs, but is not willing to stimulate
additional usage. Therefore, in this equilibrium, the competition effect fully determines the
CSP’s strategic investment choice. For example, this can be seen in Figure 6a, which illustrates
that the CSP sets its quality according to $q_0^{MP} = \delta d = 1.5$ within region IIb. In region IIc the
strategies from region IIb no longer constitute an equilibrium. Rather, the CSP chooses its
quality investment such that in Stage 2 the PDM’s reward according to its best response to
the CSP’s price exactly compensates consumers for their transaction costs and thus ensures
usage of the PDM. Finally, in region III, the CSP increases its quality investment for high
levels of the PDM’s efficiency. As the PDM rewards consumers in excess of their transaction
costs, the CSP can now also raise its price, thus leading to a quality increase due to the
appropriation effect.

**Proposition 6 (Consumers’ transaction costs)** When consumers face transactions costs
to use the PDM, the CSP can foreclose the entry of the PDM if the PDM is relatively inef-
ficient ($\rho$ is low). For intermediate efficiency levels of the PDM, a minimum-income equi-
librium emerges, in which the PDM offers a reward that is just enough to offset consumers’
transaction costs ($r = k + \epsilon / x$). For high efficiency levels of the PDM, a positive-income equi-
librium emerges, in which the PDM’s rewards exceed consumers’ transaction costs ($r > k / x$).
Propositions 1 to 4 continue to hold qualitatively for the range of the minimum-income and
positive-income equilibrium.

Remarkably, in region Ib, the CSP may reduce its service quality in anticipation of a
possibly active PDM, but, in consequence the PDM will not enter the market. Thus, the
mere threat of a PDM may already lead to a strategic quality reduction by the CSP, above
and beyond the competition and displacement effect in the actual presence of a PDM.
6.3 Data analytics technology investments and PDM

Until now we have assumed that the firms’ ability to generate revenues from data is exogenously given. However, in practice, the economic value of data sold on the market is heavily influenced by firms’ strategic decisions to process, enrich and analyze the collected raw user data. For example, a CSP can improve its efficiency on the data market by undertaking long-term investments in the development of software, which allows for additional insights from the collected data (see, e.g., the software analytics technology at Twitter (Lin and Kolcz, 2012) and LinkedIn (Sumbaly et al., 2013)), in hardware and infrastructure, which facilitate storing and processing big data (see, e.g., Müller et al., 2018), and in human resources such as data scientists (see, e.g., Tambe, 2014). Likewise, a PDM can strategically decide on the technology investments that determine the quality of its data sold to data brokers, thus determining its efficiency on the data market \( \varrho \) endogenously.

To account for endogenous investments in data analytics technology, we extend our base model by introducing a Stage 0 that precedes all other stages of the base model. In this stage both the CSP and the PDM (if present) simultaneously decide on their optimal investment levels in data analytics technology, denoted by \( t \) and \( \varrho \), respectively. Thereby, they anticipate the effect of their investment choice on subsequent stages. In line with the empirical literature on data analytics applications (Junqué de Fortuny et al., 2013; Li et al., 2016; Martens et al., 2016), we assume diminishing returns to scale, i.e., that the marginal benefit of investments decreases as a firm moves closer to the maximum economic value that can be generated from the collected data. To this end, we consider the gross revenues (excluding the competitive impact of the PDM) to be \( \sqrt{t} dx \) for the CSP and \( \sqrt{\varrho} dx \) for the PDM. While we assume that the revenues of the CSP and PDM are the same, provided both have made the same technology investment, we do allow for differences in investment costs. Specifically, we assume a standard quadratic cost function for the CSP \( C(t) = at^2 \) and the PDM \( C(\varrho) = b \varrho^2 \), parameterized by \( a \) and \( b \). For \( a < b \), we can model market situations where the CSP has a relative cost advantage over the PDM, and vice versa for \( a > b \). In particular, we will also consider the case where both firms have a symmetric cost function, i.e., \( a = b \). In the following, we characterize equilibrium market outcomes and the impact of the PDM, whereas the complete analysis is relegated to
Appendix D.

**Data analytics technology investments without PDM (A):** In the monopoly equilibrium, the CSP chooses a positive technology investment level $t^A > 0$. As in the base model, the CSP sets a a positive quality level ($q^A > 0$) and a zero price ($p^A = 0$). Clearly, in equilibrium the CSP will invest less in technology when investment costs increase ($\partial t^A/\partial a < 0$). Eventually, investments converge to zero for very high costs. Likewise, the CSP’s investments in quality decrease when investments in data analytics technology become costlier ($\partial q^A/\partial a < 0$). In consequence, lower costs for data analytics technology also benefits consumers due to the higher service quality ($\partial CS^A/\partial a < 0$). Moreover, technology and quality investments increase as data brokers’ valuation for data grows, i.e., $\partial t^A/\partial d > 0$ and $\partial q^A/\partial d > 0$.

**Data analytics technology investments with PDM (AP):** With a PDM, there are again two types of equilibria: A *minimum-income equilibrium* arises, if the CSP has a sufficiently large cost advantage over the PDM ($a < \bar{\pi}(b)$). Intuitively, this translates into a situation where the PDM is much less efficient on the data market than the CSP, i.e., $g^{AP} < t^{AP}$. In contrast, there exists a *positive-income equilibrium*, if (i) the PDM has a cost advantage over the CSP ($a > b$), (ii) technology investment costs of both firms are symmetric ($a = b$) or (iii) the PDM’s cost disadvantage is not too large ($a > \bar{\delta}(b)$).\(^5\)

Figure 7 illustrates investment and pricing decisions for these two equilibrium outcomes based on numerical examples. The two panels on the left side depict a situation, where the PDM has a large cost disadvantage vis-à-vis the CSP ($a = 0.2, b = 1$). Given this disadvantage, the PDM invests less in data analytics technology, as illustrated in Panel (a), making it less efficient on the data market ($g^{AP} < t^{AP}$). In consequence, the epsilon-reward equilibrium emerges and the PDM affects the CSP’s quality and pricing decisions solely through the *competition effect* as shown in Panel (c). Therefore, the CSP’s quality investment is strictly lower with the PDM for any $\delta < 1$, whereas the PDM’s reward and the CSP’s price are close to zero ($p^{AP} = r^{AP} = \epsilon \approx 0$) for all levels of $\delta$. In contrast, if the PDM has a cost advantage over the CSP (as illustrated in the panels on the right side of Figure 7 for the case of $a = 2, b = 0.1$),

\(^5\)See Appendix D for a derivation of the cost thresholds.
the PDM invests more in data analytics technology than the CSP and will be more efficient on the data market (see Panel (b)). Therefore, the PDM offers a positive reward in order to stimulate the generation of more data, i.e., \( r^{AP} > \epsilon \). If the relative cost advantage of the PDM is large (\( a < b \)) and the competitive intensity on the data market is low (\( \delta \) is high), the appropriation effect can outweigh the competition and the displacement effect, giving the CSP an incentive to increase its quality investment above the level without a PDM. For example, in Panel (d) this is the case for high values of \( \delta \) that exceed the threshold \( \delta^* = 0.913 \). Thus, qualitatively the same market outcomes as in the base model arise.

Consequently, also the insights from Propositions 1 to 4 continue to hold qualitatively. This includes that consumers truly benefit from the presence of a PDM, i.e., \( CS^{AP} > CS^A \)
only if the PDM has a sufficiently large cost advantage and is therefore willing to pay a high reward. Otherwise consumers are worse off. In particular, in the special case, where the CSP and the PDM have identical investment costs, consumers are never better off with a PDM, i.e., \( CS^{AP} < CS^A \) for \( a = b \). Although the PDM is willing to pay a positive reward in this situation, the displacement effect already outweighs the appropriation effect, even without any competitive effect (\( \delta = 1 \)).

**Proposition 7 (Data analytics technology)** With endogenous investments in data analytics technology, the insights from Proposition 1 to 4 continue to hold qualitatively. If the PDM has a large cost disadvantage, the minimum-income equilibrium emerges. If the PDM’s cost disadvantage is small or if the PDM has an advantage over the CSP, the positive-income equilibrium arises. Moreover, consumers are only better off with a PDM if the PDM has a sufficiently large cost advantage over the CSP.

7 Conclusion

In this study, we focus on the impact of a personal data market on the strategic decisions of online content and service providers, such as how much providers should invest into the quality of their service and how they should optimally monetize their service. In practice, most content and service providers collect user data and rely on revenues from this data to finance their service. However, providers may also charge consumers a price for using their service. Our results show that a personal data market has important strategic implications beyond a direct monetary benefit for consumers. In fact, the presence of a personal data market together with consumers’ ability to transfer data has several significant effects on a content and service provider’s optimal strategic actions and thus on market outcomes in the data economy.

Specifically, we identify three main economic effects that arise in the context of a personal data market. The relative strength of those effects determine content and service providers’ investment incentives and ultimately they determine whether consumers benefit from a personal data market or not. First, a personal data market may free ride on the quality investments
of a content and service provider. If the affected provider suffers from reduced excludability of its collected data through lower revenues on the data market, it will lower its investment in quality (competition effect). Second, a personal data market may stimulate consumption and data creation by remunerating consumers through a positive reward for their data. As monetary remuneration stimulates consumers’ usage, the content and service provider has an additional incentive to decrease its quality investment (displacement effect). Third, remuneration increases consumers’ surplus and thus also raises the content and service provider’s incentive to charge a higher own price. This, in turn, has a positive impact on the content and service provider’s investment in quality (appropriation effect).

Due to the displacement effect, a content and service provider has an incentive to decrease its quality investment even beyond the reduction caused by the competition effect. However, the incentives to reduce quality may be countervailed by the appropriation effect. All three effects can have a significant impact on a content and service provider’s strategic price and quality choices, its profit and the impact of personal data markets on consumer surplus. As shown in our analysis the relative strength of these effects is primarily determined by the personal data market’s efficiency on the data market, i.e., its ability to generate data revenues relative to the content and service provider. Most notably, the personal data market can make consumers only better off if it has an actual efficiency advantage over the content and service provider. This result is robust across a number of model extensions.

7.1 Managerial and Policy Implications

Our analysis bears several important implications for managers and policy makers that are concerned with markets, where data constitutes an important economic good.

The impact of a personal data market on a content and service provider’s profit is mainly determined by the competitive relationship of both firms on the data market and their relative ability to generate revenues from data. If the personal data market sells its data to the same or identical data brokers as the content and service provider and does not significantly alter, process or enhance the data, the competitive effect will likely be large and thus the content and service provider will likely be worse off. In contrast, if the personal data market is able to
enrich the original data set, e.g., because consumers can add additional data points, update outdated information or remove incorrect data records, the content and service provider is likely to benefit from a personal data market. In this case the personal data market subsidizes consumers’ usage and therefore benefits the content and service provider.

In today’s Internet ecosystem, many content and service providers follow a business model where they offer consumers their services for a zero price. Instead of direct monetization in the consumer market, providers rely on the collection of user data, which they then sell on the data market. Any positive price for consumers would likely hamper the ability to collect data, because a zero price stimulates usage intensity. Our results indicate that personal data markets could have a significant impact on this prototypical business model of the Internet, if they are indeed able to generate sufficiently high revenues on the data market.

For moderate levels of efficiency, i.e., if the personal data market’s efficiency disadvantage on the data market relative to the content and service provider is not too large, positive remuneration of consumers by the personal data market may make it optimal for the content and service provider to reduce quality investment, but, at the same time, to stimulate usage through direct reward schemes for consumers. For example, providers may offer consumers vouchers or free products. This may be profitable with a personal data market, because the content and service provider saves costly investments in quality, but the burden to stimulate usage through direct compensation is shared with the personal data market.

In contrast, if the personal data market has a sufficiently large efficiency advantage over a content and service provider, the provider may start to charge consumers for their usage of its service. Instead of solely relying on indirect monetization on the data market, the content and service provider may now directly benefit from the existence of the personal data market by appropriating a share of the reward paid out in the consumer market. Because remuneration by the personal data market increases consumers’ surplus, they are more likely to accept a (higher) price for consuming Internet services and content without lowering their usage. The content and service provider may now even consider to increase the quality of its service, because it can appropriate the generated surplus through the positive price.

Regulators and policy makers who are interested in maximizing consumer surplus should
carefully monitor the reward that a personal data market offers consumers for their data. In general, they should be aware that the positive effect of user remuneration can be offset by the negative effects of a content and service provider’s strategic reduction of quality investments. In particular, if the personal data market is only willing to pay marginal rewards, which barely ensure the transfer of consumers’ data, it is likely that consumers are ultimately worse off due to a lower quality of the available content and services. Only if the personal data market has a significant competitive advantage over the content and service provider and at the same time intensity of competition between both firms is not too strong, consumers are likely to benefit from a personal data market. It is in this case, however, that we expect content and services providers to raise prices compared to the status quo. In terms of total welfare this is likely to be beneficial, because the price instrument allows content and service providers to benefit from a personal data market and also gives providers an incentive to invest in the quality of their service.

7.2 Limitations and Future Work

In order to focus on the main strategic effects that arise from the emergence of personal data markets, we have abstracted from the fact that users may supply user data of different economic value. In particular, an extrinsic price mechanism as institutionalized by personal data markets may incentivize consumers to create fake data in order to artificially increase the size of the data set that they sell. On the other hand, proponents of personal data markets have argued that the direct participation of the consumers in the data transaction increases transparency and trust of those consumers and may therefore lead to more accurate and detailed data provisions. Future studies should explicitly model and consider these trade-offs concerning data quality in the context of personal data markets. While we expect that this will provide more nuanced results, the main strategic effects identified here would continue to exist. Moreover, our theoretical analyses may guide empirical studies when they assess the economic impact of data portability and personal data markets. In particular, the effects identified and delineated in this study may provide a clear conceptual framework for measuring the overall effects of those institutions.
References


European Data Protection Supervisor (2016). EDPS opinion on personal information management systems: Towards more user empowerment in managing and processing per-


Appendix

A Base model: Equilibrium existence

We identify two types of equilibria: the positive-income equilibrium, which is derived as the interior solution, and the minimum-income equilibrium, which is derived as a corner solution. Note that if consumers could not exit the PDM, the personal data market would be able to set a reward \( r < \epsilon \), which is an interior solution for low levels of efficiency \( \varrho \) of the PDM. However, in this case, consumers would leave the PDM and thus the PDM is restricted to set a positive reward \( r \geq \epsilon \). In these cases the corner solution constitutes an equilibrium.

We denote values of the minimum-income equilibrium by the subscript \( \epsilon \), and positive-income equilibrium values by the subscript +:

- **Positive-income equilibrium:**
  \[
  q^+_{\text{MP}} = \frac{2}{7} \delta d(1 + \varrho), \quad p^+_{\text{MP}} = \frac{1}{7} \delta d(3\varrho - 4), \quad r^+_{\text{MP}} = \frac{1}{7} \delta d(4\varrho - 3), \quad x^+_{\text{MP}} = \frac{3}{7} \delta d(1 + \varrho),
  \]
  \[
  \Pi^+_{\text{MP}} = \frac{1}{7} \delta^2 d^2 (1 + \varrho)^2, \quad \Gamma^+_{\text{MP}} = \frac{9}{49} (\delta d + \delta \varrho d)^2.
  \]

- **Minimum-income equilibrium:**
  \[
  q^\epsilon_{\text{MP}} = \delta d + \epsilon, \quad p^\epsilon_{\text{MP}} = \epsilon, \quad r^\epsilon_{\text{MP}} = \epsilon, \quad x^\epsilon_{\text{MP}} = \delta d + \epsilon,
  \]
  \[
  \Pi^\epsilon_{\text{MP}} = \frac{1}{2} (\delta d + \epsilon)^2, \quad \Gamma^\epsilon_{\text{MP}} = (\delta d + \epsilon)(\delta \varrho d - \epsilon).
  \]

In order to prove the existence of the equilibrium, we will first assume that the two types of equilibrium candidates coexist for all \( \varrho \in (0, \infty) \). As the CSP has a first mover advantage over the PDM and \( \frac{\partial r^\text{MP}}{\partial q} < 0 \), it can determine the type of equilibrium by its quality decision. Thus, we will, secondly, identify the intervals in which the CSP is willing to choose the minimum-income equilibrium and the positive-income equilibrium. We will show that the equilibrium candidates might exist throughout the respective intervals, which completes the proof.
At first, we identify the level of $\varrho$ for which the CSP is indifferent between the two types of equilibrium candidates, i.e. $\Pi^* = \Pi^*(\varrho)$. Straightforward calculations give $\varrho = \sqrt{\frac{7}{2\sqrt{2}} \frac{d^2\varphi + 2d\varphi + \varphi}{d^2\varphi} - 1}$. For $\delta d > 0$, we are able to then make use of $\Pi_{MP}^* \geq 0$, the fact that $\Pi_{MP}^*$ does not depend $\varrho$, and $\frac{\partial \Pi_{MP}(\varrho)}{\partial \varrho} > 0$ to derive that the CSP decides in favor of the minimum-income outcome if $\varrho \leq \overline{\varrho}$ and in favor of the positive-income outcome if $\varrho \geq \overline{\varrho}$.

We proceed by proving the existence of these types of equilibrium candidates in the respective intervals and start with the positive-income equilibrium candidate: For $\varrho > \overline{\varrho}$, the CSP adjusts its quality investments in order to induce the positive-income equilibrium. The equilibrium might exist as long as user participation is ensured, i.e. if $r_{MP} + (q_{MP} + \varrho) > \epsilon$. This is the case if $\varrho > \overline{\varrho} = \frac{3d\delta + 7\epsilon}{4d\delta}$.

We are also interested in the range of $\varrho$, for which there might be a minimum-income equilibrium. In order for the users to transfer data to the PDM, the PDM must at least set $r \geq \epsilon$. Whenever $\delta d \varrho > \epsilon$, the PDM at least wants to ensure user participation. This means that the market participation of a PDM cannot be influenced by the CSP. Thus, as long as $r_{MP}^* (q_{MP}) \leq \epsilon$, the PDM will set $r = \epsilon$. This is the case for $\varrho \leq \overline{\varrho} = \frac{d\delta + 2\epsilon}{4d\delta}$.

To complete the proof, it is easy to see that $\overline{\varrho} < \overline{\varrho} < \overline{\varrho}$. Therefore, the types of equilibria exist in the relevant ranges. $\square$

### B Competing CSPs and PDM

In the extension with competing CSPs, we index all outcomes in the scenario without personal data market with $D$ and alls outcomes in the scenario with personal data market with $DP$.

**Duopolistic CSPs without PDM:** In Stage 3, consumers’ utility is maximized by choosing a usage level of

$$x_i^D(q_i, q_j, p_i, p_j) = \frac{q_i - p_i - \gamma (q_j - p_j)}{1 - \gamma^2},$$

where $i = A, B$ and $j$ denotes the respective CSP rival.

In Stage 2, anticipating the usage levels, the CSPs simultaneously choose prices to maximize their profits in this stage, with Stage 2 profits (quality investments being sunk) given
by
\[
\pi_{i,2}(p_i, x_i^D) = (d + p_i) x_i^D = (d + p_i) \frac{q_i - p_i - \gamma(q_i - p_j)}{1 - \gamma^2}.
\]
Solving first-order conditions \( \frac{\partial \pi_{i,2}}{\partial p_i} = 0 \) and \( \frac{\partial \pi_{i,2}}{\partial x_i} = 0 \) simultaneously, we obtain optimal Stage 2 prices:
\[
p_i^D(q_i, q_j) = \frac{(2 - \gamma^2)q_i - (2 + \gamma)d - \gamma q_j}{4 - \gamma^2}.
\]
In Stage 1, the CSPs simultaneously decide about their quality level by maximizing their profits in this stage, with Stage 1 profits given by
\[
\pi_{i,1}(q_i, p_i^D, p_j^D, x_i^D) = (d + p_i^D) x_i^D(p_i^D, p_j^D) - \frac{1}{2} q_i^2.
\]
This yields candidate equilibrium quality levels of
\[
q_i^D = \frac{2(2 - \gamma^2) d}{(2 - \gamma)(1 - \gamma)(2 + \gamma) + 4}.
\]
The candidate subgame-perfect Nash-equilibrium is then characterized by the following strategies and outcomes:
\[
q_i^D = \frac{2(2 - \gamma^2) d}{(2 - \gamma)(1 - \gamma)(2 + \gamma) + 4}, \quad p_i^D = -\frac{(1 - \gamma)(\gamma + 4) d\gamma}{(2 - \gamma)(1 - \gamma)(2 + \gamma) + 4},
\]
\[
x_i^D = \frac{4(2 - \gamma^2) d}{(2 - \gamma)(1 - \gamma)(2 + \gamma) + 4},
\]
\[
\pi_i^D = \frac{8 - \gamma^6 + 7\gamma^4 - 16\gamma^2}{((2 - \gamma)(1 - \gamma)(2 + \gamma) + 4)^2} d^2, \quad CS_i^D = \frac{(\gamma + 1)(4 - \gamma^2)^2 d^2}{((2 - \gamma)(1 - \gamma)(2 + \gamma) + 4)^2}.
\]
For \( \gamma > 0.827891 \equiv \tau \) this strategy yields negative profits for both CSPs, which cannot constitute an equilibrium outcome. Clearly, a CSP could do better if it would set \( q_i^D = 0 \) instead, because thereby it can always secure a profit of at least zero.

For \( \gamma > \tau \), it is easy to see that the reaction function in Stage 1 \( q_i^{D,r}(q_j) \) to a zero quality competitor is negative. Therefore, CSP \( i \) would like to set a negative quality in response to CSP \( j \) opting for a zero quality level. As CSP \( i \)'s quality is bounded below by
\( q_i \geq 0 \), and it, thus, cannot further reduce its quality, it does best by playing the minimum feasible level of \( q_i = 0 \). Due to symmetry of the CSP’s and \( \pi^D_i(q_i = q_j = 0) = \pi^D_j(q_i = q_j = 0) \geq 0 \) for \( \gamma > \gamma \), there is an equilibrium specified by \( q_i = q_j = 0 \), \( p_{i,0}^D = p_{j,0}^D = -\frac{(2+\gamma)d}{4-\gamma^2} \) and \( x_{i,0}^D = x_{j,0}^D \).

**Duopolistic CSPs with PDM:** In Stage 3, consumers’ optimal usage level at CSP \( i \) is given by

\[
x_{i}^{DP}(q_i, q_j, p_i, p_j, r) = \frac{q_i - p_i + r(1 - \gamma) - \gamma(q_j - p_j)}{1 - \gamma^2},
\]

In Stage 2, the PDM and the CSPs simultaneously choose their prices and rewards by maximizing their Stage 2 profits (quality investments being sunk) given by

\[
\Gamma^{DP}(r, x_i^{DP}, x_j^{DP}) = (d\delta q - r)(x_i^{DP} + x_j^{DP}) = (d\delta q - r)\frac{(q_i + q_j - p_i - p_j + 2r)}{1 + \gamma},
\]

\[
\pi^{DP}_{i,2}(p_i, x_i^{DP}) = (d\delta + p_i)x_i^{DP} = (d\delta + p_i)\frac{r(1 - \gamma) + q_i - p_i - \gamma(q_j - p_j)}{1 - \gamma^2},
\]

where, like in the base model, \( \delta \in [0, 1] \) captures the (negative) competitive effect of the PDM on CSPs’ profit on the data market relative to the baseline without PDM. In Stage 2, restricting attention to the parameter range \( \gamma \leq \gamma \), equilibrium prices and rewards are characterized by

\[
r^{DP}(q_i, q_j) = \begin{cases} \frac{2d\delta((2-\gamma)\rho-1)-q_i-q_j}{2(\beta-\gamma)} & \text{if } q_i + q_j \leq \overline{q}^{DP} := 2d\delta((2 - \gamma)\rho - 1), \\ 0 & \text{otherwise,} \end{cases}
\]

\[
p^{DP}_{i}(q_i, q_j) = \begin{cases} \frac{2(\gamma+2)d\delta((1-\gamma)\rho-2)-\gamma(2\gamma-1)-5q_i-3\gamma q_j-q_j}{4(\beta-\gamma)(\gamma+2)} & \text{if } q_i + q_j \leq \overline{q}^{DP}, \\ \frac{(2-\gamma^2)q_i-\gamma q_j-(\gamma+2)d\delta}{4-\gamma^2} & \text{otherwise.} \end{cases}
\]

In Stage 1, CSPs maximize their profits by simultaneously choosing their quality levels. Like in the base model, two types of equilibria exist, depending on the PDM’s efficiency.
level \( \varrho \):

\[
q_i^{DP} = \begin{cases} 
\frac{\gamma(2\gamma-1)-5}{(2-\gamma)(\gamma-1)(\gamma+2)-13} & \text{if } \varrho \geq \varrho^{DP}, \\
2\gamma^2-2 & \text{if } \varrho \leq \varrho^{DP}.
\end{cases}
\]

In the following, we will assume \( \epsilon = 0 \). The zero-reward equilibrium (which then corresponds to the minimum-income equilibrium) exists for \( \varrho \in [0, \varrho^{DP}] \) and is characterized by:

\[
q_i^{DP}, 0 = \frac{2(2-\gamma^2) d \delta}{(2-\gamma)(1-\gamma)\gamma(\gamma+2) + 4},
\]

\[
p_i^{DP}, 0 = \frac{-\gamma((1-\gamma)\gamma + 4d \delta}{(2-\gamma)(1-\gamma)\gamma(\gamma+2) + 4},
\]

\[
r_i^{DP}, 0 = 0,
\]

\[
x_i^{DP}, 0 = \frac{(4-\gamma^2) d}{(2-\gamma)(1-\gamma)\gamma(\gamma+2) + 4},
\]

\[
\pi_i^{DP}, 0 = \frac{8 - 16\gamma^2 + 7\gamma^4 - \gamma^6}{((2-\gamma)(1-\gamma)\gamma(\gamma+2) + 4)^2} d^2 \delta^2,
\]

\[
\Gamma^{DP}, 0 = \frac{(4-\gamma^2) d^2 \delta^2 d \rho}{(2-\gamma)(1-\gamma)\gamma(\gamma+2) + 4},
\]

\[
CS^{DP}, 0 = \frac{(\gamma + 1)(4-\gamma^2) d^2 \delta^2}{((2-\gamma)(1-\gamma)\gamma(\gamma+2) + 4)^2}.
\]
The positive-income equilibrium exists for $\varrho \in [\varrho_{DP}, \infty)$ and is characterized by

$$q_{i,t}^{DP} = \frac{(\gamma(1 - 2\gamma) + 5)d\delta(\rho + 1)}{(2 - \gamma)\gamma((1 - \gamma)\gamma + 7) + 13},$$

$$p_{i,t}^{DP} = \frac{d\delta((\gamma - 1)\gamma - 7)\gamma\rho + 2(\gamma - 1)\gamma + \rho - 13) + 6\rho - 7)}{(2 - \gamma)\gamma((1 - \gamma)\gamma + 7) + 13},$$

$$\gamma_{i,t}^{DP} = \frac{d\delta(\gamma^3 + (\gamma((\gamma - 2)\gamma - 5) + 7)\rho - 7\gamma - 6)}{(2 - \gamma)\gamma((1 - \gamma)\gamma + 7) + 13},$$

$$x_{i,t}^{DP} = \frac{(3 - \gamma)(\gamma + 2)d\delta(\rho + 1)}{(2 - \gamma)\gamma((1 - \gamma)\gamma + 7) + 13},$$

$$\pi_{i,t}^{DP} = \frac{(47 - \gamma(\gamma(2\gamma((\gamma - 2)\gamma - 10) + 12) + 75) - 14))d^2\delta^2(\rho + 1)^2}{2((2 - \gamma)\gamma((1 - \gamma)\gamma + 7) + 13)^2},$$

$$\Gamma_{+}^{DP} = \frac{2(\gamma + 1)(-\gamma^2 + \gamma + 6)^2 d^2\delta^2(\rho + 1)^2}{((2 - \gamma)\gamma((1 - \gamma)\gamma + 7) + 13)^2},$$

$$CS_{+}^{DP} = \frac{(\gamma + 1)(-\gamma^2 + \gamma + 6)^2 d^2\delta^2(\rho + 1)^2}{((2 - \gamma)\gamma((1 - \gamma)\gamma + 7) + 13)^2}.$$

### C Transaction costs for using the PDM

**Equilibrium characterization:** We derive the equilibrium outcome by backward induction. In Stage 3, consumers choose over their optimal usage level and whether to visit the PDM:

$$x^{TP}(q,p,r) = \begin{cases} 
q - p + r & \text{if } r(q - p + r) \geq k, \\
q - p & \text{if } r(q - p + r) < k.
\end{cases}$$

Consumers visit the PDM if $rx \geq k$, i.e., if their reward for selling the data is greater than the transaction costs for using the PDM. We will refer to this condition later on as the consumer participation constraint.

In Stage 2, the CSP and the PDM simultaneously choose price $p$ and reward $r$ by maxi-
mizing their Stage 2 profits given by

\[ \Gamma^T_P(r, x^T_P) = \begin{cases} 
(d\delta - r)x^T_P & \text{if } rx^T_P \geq k, \\
0 & \text{otherwise}. 
\end{cases} \]

\[ \pi^T_P(p, x^T_P) = \begin{cases} 
(d\delta + p)x^T_P & \text{if } rx^T_P \geq k, \\
(d + p)x^T_P & \text{otherwise}. 
\end{cases} \]

In addition to the consumer participation constraint, the PDM must make positive profits in equilibrium. That is, a PDM’s reward must not exceed its maximum return per data from the data market, i.e. \( r \leq r_{\text{max}} = d\delta \). In the following we denote this as the *profitability constraint*.

Given the consumer participation constraint and the profitability constraint, there are six equilibrium candidates at Stage 2, which we will specify with their respective feasibility conditions. These equilibrium candidates can be subsumed in three categories: Two foreclosure equilibrium candidates (Ia, Ib), where there is no PDM in the market, three minimum-income equilibrium candidates (IIa, IIb, IIc), where the PDM is in the market but user remuneration does not exceed the transaction costs, i.e. \( rx = k \), and one positive-income equilibrium candidate (III) which refers to the unconstrained best response equilibrium in Stage 2.

In equilibrium candidates (Ia) and (Ib) the PDM offers the maximum reward of \( r_{\text{max}} \), but does not attract users. In (Ia) the CSP’s monopoly strategy is sufficient to keep consumers from using the PDM. In (Ib) the CSP chooses its price \( p \) such that the participation constraint is not fulfilled. We assume that in this equilibrium candidate consumers abstain from using the PDM in the case of indifference. In contrast, in equilibrium candidates (IIa), (IIb) and (IIc) we assume that consumers visit the PDM in case of indifference. In practice, a CSP can ensure user participation (non-participation) at the PDM by slightly decreasing (increasing) its price by \( \epsilon \). The foreclosure equilibrium candidates are specified by:
Candidate (Ia) for $0 < q \leq \frac{2k-2\delta^2 d^2 \rho^2 - \delta^2 d^2 \varrho^2}{d \varrho}$:

$$r_{Ta}^{TP}(q) = d \varrho,$$

$$p_{Ta}^{TP}(q) = \frac{q - d}{2}.$$

Candidate (Ib) for $0 < q$:

$$r_{Tb}^{TP}(q) = d \varrho,$$

$$p_{Tb}^{TP}(q) = -\frac{k - d^2 \delta^2 \varrho^2 - d \delta q \varrho}{d \varrho}.$$

In the minimum-income candidates (IIa), (IIb) and (IIc) consumers transfer their data to the PDM. In candidate (IIa) the participation constraint and the profitability constraint bind with equality. This means, the CSP chooses $p$ in order to profit from a maximum possible usage subsidization of the PDM. In candidate (IIb) only the participation constraint binds with equality. Here, the PDM’s optimal reward in response to a CSP’s quality would not be sufficient to attract consumers given their transaction costs. Thus, the CSP induces a reward that is larger than the PDM’s unconstrained reward, but lower than the maximum feasible reward $r_{max}$. In equilibrium candidate (IIc) the CSP adjusts its price such that the PDM’s reaction function to this price is sufficient to keep the PDM in the market. The minimum-income equilibrium candidates are formally specified by:

Candidate (IIa) for $0 < q$:

$$r_{IIa}^{TP}(q) = d \varrho,$$

$$p_{IIa}^{TP}(q) = -\frac{k - d^2 \delta^2 \varrho^2 - d \delta q \varrho}{d \varrho}.$$
Candidate (IIb) for Max\([0; \sqrt{d^2 \delta^2 \varrho^2 - 4k + d\delta \varrho - d\delta; \frac{2k-d^2\delta^2\varrho}{d\delta\varrho}}] < q:\]

\[
r_{\text{IIb}}(q) = \frac{2k}{d\delta + q},
\]

\[
p_{\text{IIb}}(q) = \frac{4k + q^2 - d^2\delta^2}{2(d\delta + q)}.
\]

Candidate (IIc) for \(0 < q\) and \(k \leq \frac{1}{4}(d\delta\varrho)^2\):

\[
r_{\text{IIc}}(q) = \frac{1}{4}\left(\sqrt{9\varrho^2 - 16k + 3\varrho}\right),
\]

\[
p_{\text{IIc}}(q) = \frac{1}{2}\left(\sqrt{9\varrho^2 - 16k + 2\varrho}\right).
\]

Finally, the positive-income equilibrium candidate (III) refers to the interior solution, in which the combination of \(q, r\) and \(p\) is such that neither the participation constraint, not the profitability constraint is binding.

Candidate (III) for Max\([0; d\delta(2\varrho - 1); \frac{1}{2}\left(d\delta(\varrho - 2) - 3\sqrt{d^2\delta^2\varrho^2 - 4k}\right)] \leq q \leq \frac{1}{2}\left(3\sqrt{d^2\delta^2\varrho^2 - 4k} + d\delta(\varrho - 2)\right)\):

\[
r_{\text{III}}(q) = \frac{1}{3}(d\delta(2\varrho - 1) - q),
\]

\[
p_{\text{III}}(q) = \frac{1}{3}(d\delta(\varrho - 2) + q).
\]

In Stage 1, the equilibrium qualities corresponding to the six aforementioned equilibrium candidates are obtained by maximizing the following respective Stage 1 profits:
\[ \pi_{T_{Pa},1}(q) = \frac{1}{4} \left( d^2 + 2dq - q^2 \right), \]
\[ \pi_{T_{Pb},1}(q) = \frac{2d\delta k \varrho (2d\delta \varrho + d + q) - d^2 \delta^2 \varrho^2 \left( 2d\delta \varrho \delta \varrho + 1 \right) + 2d\delta dq \varrho + q^2) - 2k^2}{2d^2 \delta^2 \varrho^2}, \]
\[ \pi_{T_{Pa},1}(q) = \frac{k(d\delta \varrho (d\delta \varrho + q) - k)}{d^2 \delta^2 \varrho^2} - \frac{q^2}{2}, \]
\[ \pi_{T_{Pb},1}(q) = \frac{1}{4} \left( (d\delta + q)^2 + 4k - 2q^2 \right), \]
\[ \pi_{T_{Pa},1}(q) = \left( \sqrt{d^2 \delta^2 \varrho^2 - 4k + d\delta + q} \right) \left( r - \sqrt{d^2 \delta^2 \varrho^2 - 4k} \right) - \frac{q^2}{2}, \]
\[ \pi_{T_{Pb},1}(q) = \frac{1}{18} \left( 2d^2 \delta^2 (\varrho + 1)^2 + 4d\delta (\varrho + 1)q - 7q^2 \right) \]

The resulting quality levels for the six equilibrium candidates are:

\[ q_{T_{Pa}} = d, \]
\[ q_{T_{Pb}} = \frac{k - d^2 \delta^2 \varrho^2}{d\delta \varrho}, \]
\[ q_{T_{Pa},a} = \frac{k}{d\delta \varrho}, \]
\[ q_{T_{Pa},b} = d\delta, \]
\[ q_{T_{Pa},c} = \frac{1}{4} \left( 3\varrho - \sqrt{9\varrho^2 - 16k} \right), \]
\[ q_{T_{Pa},d} = \frac{2}{7} \delta d \left( 1 + \varrho \right) \]

Comparing profits from the feasible candidates yields the profit maximizing quality level for each parameter configuration.
**Numerical Example:** Not all equilibrium types exist for any set of parameters $k, \delta, d$ and $v$. For illustrative reasons we will use throughout the main body a specific example ($k = 3$, $\delta = 0.5$, $d = 3$ and $v = 3$) for which all equilibrium candidates emerge as an equilibrium outcome for a particular parameter region along $\varrho$. For this example, equilibrium quality is specified by:

$$q_{TP} = \begin{cases} 
  3 & \text{if } \varrho < \sqrt[3]{\frac{7}{3}} - 1, \\
  \frac{4 - 3\varrho^2}{2\varrho} & \text{if } \sqrt[3]{\frac{7}{3}} - 1 < \varrho < 0.75, \\
  \frac{2}{\varrho} & \text{if } 0.75 < \varrho < \frac{4}{3}, \\
  \frac{3}{2} & \text{if } \frac{4}{3} < \varrho < \frac{7}{3}, \\
  \frac{1}{4} \left( 3\varrho - \sqrt{9\varrho^2 - 48} \right) & \text{if } \frac{7}{3} < \varrho < 2.55, \\
  \frac{3(\varrho + 1)}{4} & \text{if } 2.55 < \varrho.
\end{cases}$$
Equilibrium prices, rewards and usage intensity are then given by

\[ p^{TP} = \begin{cases} 
0 & \text{if } \varrho < 0.75, \\
\frac{3\varrho}{2} & \text{if } 0.75 < \varrho < \frac{4}{3}, \\
2 & \text{if } \frac{4}{3} < \varrho < \frac{7}{3}, \\
\frac{1}{4}(\sqrt{9\varrho^2 - 48} + 3\varrho) & \text{if } \frac{7}{3} < \varrho < 2.55, \\
\frac{3}{14}(3\varrho - 4) & \text{if } 2.55 < \varrho,
\end{cases} \]

\[ r^{TP} = \begin{cases} 
1.5\varrho & \text{if } \varrho < \frac{4}{3}, \\
2 & \text{if } \frac{4}{3} < \varrho < \frac{7}{3}, \\
\frac{1}{4}(\sqrt{9\varrho^2 - 48} + 3\varrho) & \text{if } \frac{7}{3} < \varrho < 2.55, \\
\frac{3}{14}(4\varrho - 3) & \text{if } 2.55 < \varrho,
\end{cases} \]

\[ x^{TP} = \begin{cases} 
3 & \text{if } \varrho < \sqrt{\frac{7}{3}} - 1, \\
\frac{4 - 3\varrho^2}{2\varrho} & \text{if } \sqrt{\frac{7}{3}} - 1 < \varrho < 0.75, \\
\frac{2}{\varrho} & \text{if } 0.75 < \varrho < \frac{4}{3}, \\
\frac{3}{2} & \text{if } \frac{4}{3} < \varrho < \frac{7}{3}, \\
\frac{1}{4}(3\varrho - \sqrt{9\varrho^2 - 48}) & \text{if } \frac{7}{3} < \varrho < 2.55, \\
\frac{9(\varrho + 1)}{14} & \text{if } 2.55 < \varrho.
\end{cases} \]
Moreover, in equilibrium, the CSP’s profit $\pi$ and the PDM’s profit $\Gamma$ are specified by:

$$\pi^{TP} = \begin{cases} \frac{9}{2} & \text{if } \varrho < \sqrt{\frac{7}{3}} - 1, \\ \frac{(-3\varrho^2 + 4)(3\varrho^2(\varrho + 4) - 4)}{8\varrho^2} & \text{if } \sqrt{\frac{7}{3}} - 1 < \varrho < 0.75, \\ \frac{3\varrho(\varrho + 1) - 2}{\varrho^2} & \text{if } \sim 0.75 < \varrho < \frac{4}{3}, \\ \frac{9}{8} & \text{if } \frac{4}{3} < \varrho < \frac{7}{3}, \\ \frac{3}{32} \left(3\varrho - \sqrt{9\varrho^2 - 48}\right) \left(\sqrt{9\varrho^2 - 48} + 4\right) & \text{if } \frac{7}{3} < \varrho < 2.55, \\ \frac{9}{25} (\varrho + 1)^2 & \text{if } \sim 2.55 < \varrho. \end{cases}$$

$$\Gamma^{TP} = \begin{cases} 0 & \text{if } 0 < \varrho < \frac{4}{3}, \\ \frac{9\varrho}{4} - 3 & \text{if } \frac{4}{3} < \varrho < \frac{7}{3}, \\ \frac{1}{16} \left(\sqrt{9\varrho^2 - 48} - 3\varrho\right)^2 & \text{if } \frac{7}{3} < \varrho < 2.55, \\ \frac{81}{192} (\varrho + 1)^2 & \text{if } \sim 2.55 < \varrho. \end{cases}$$

Finally, consumer surplus in equilibrium is specified by:

$$CS^{TP} = \begin{cases} \frac{15}{2} & \text{if } \varrho < \sqrt{\frac{7}{3}} - 1, \\ \frac{9\varrho^2}{8} + \frac{2}{\varrho^2} & \text{if } \sqrt{\frac{7}{3}} - 1 < \varrho < 0.75, \\ \frac{2}{\varrho^2} & \text{if } \sim 0.75 < \varrho < \frac{4}{3}, \\ \frac{9}{8} & \text{if } \frac{4}{3} < \varrho < \frac{7}{3}, \\ \frac{3}{16} \left(\varrho \left(3\varrho - \sqrt{9\varrho^2 - 48}\right) - 8\right) & \text{if } \frac{7}{3} < \varrho < 2.55, \\ \frac{81}{382} (\varrho + 1)^2 & \text{if } \sim 2.55 < \varrho. \end{cases}$$

Figure A1 depicts the CSP’s price, the PDM’s reward, the PDM’s profit and the consumer surplus for this numerical example.
Figure A1: Firms’ equilibrium strategies, PDM’s profit and consumer surplus (for $k = 3, \delta = 0.5, d = 3, v = 3$).

**D Data analytics technology investments**

In the extension with data analytics technology, we index all outcomes in the scenario without personal data market with $A$ and all outcomes in the scenario with personal data market with $AP$. To ease notation, we assume $\epsilon \rightarrow 0$.

**D.1 Scenario without Personal Data Market**

In the scenario without a PDM, the CSP’s profit function is given by

$$\Pi^A(x, p, q, t) = \sqrt{t} dx + px - \frac{q^2}{2} - \alpha t^2.$$

The CSP chooses its optimal investment in data analytics technology $t$ in Stage 0, with all later stages being the same as in the base model. Moreover, consumers’ utility function is the
same as in the base model without a PDM, i.e. $U^A := U^M$.

**Equilibrium derivation**  In Stage 3 consumers choose their optimal level of usage $x^A(p,q) = q - p$. In Stage 2 investment costs with respect to quality and data analytics technology are already sunk. Anticipating consumers’ usage decision, the CSP chooses its optimal price, given $t$ and $q$, $p^A(q,t) = \frac{q - \sqrt{t}}{2}$. In Stage 1, the CSP chooses the quality level that maximizes its profit, given its investment in data analytics technology $t$ in the preceding stage: $q^A(t) = d\sqrt{t}$.

It is easy to see that quality investment increases with investment in data analytics technology, i.e. $\frac{\partial q^A(t)}{\partial t} > 0$. In Stage 0, anticipating all subsequent decisions, the CSP chooses its optimal investment in data analytics technology:

$$t^A = \arg \max_{t \in \mathbb{R}_0^+} \Pi^A(x^A, p^A, q^A, t) = \frac{d^4}{4a}.$$ 

**Monopoly equilibrium** values are then given by:

$$t^A = \frac{d^2}{4a}, \quad q^A = \frac{d^2}{2\sqrt{a}}, \quad p^A = 0, \quad x^A = \frac{d^4}{2\sqrt{a}},$$

$$\Pi^A = \frac{d^4}{16a}, \quad CS^A = \frac{d^4}{8a}.$$ 

**D.2 Scenario with Personal Data Market**

In the scenario with a PDM, the profit functions of the CSP and the PDM are given by:

$$\Pi^{AP}(x,p,q,t) = \sqrt{t} \delta dx + px - \frac{q^2}{2} - at^2,$$

$$\Gamma^{AP}(x,r,\varrho) = \sqrt{\varrho} \delta dx - rx - b\varrho^2.$$ 

The CSP and PDM choose their optimal level of investment in data analytics technology independently, but simultaneously in Stage 0. Stages 1 to 3 are then again the same as in the base model with PDM (see Section 4.1). Moreover, consumers’ utility function is the same as in the base model with a PDM, i.e., $U^{AP} := U^{MP}$. 

54
Equilibrium derivation: In Stage 3 consumers choose their optimal usage taking into account quality and price of the CSP as well as the reward offered by the PDM: \( x^{AP}(p, q, r) = q - p + r \). In Stage 2 the CSP and the PDM simultaneously set \( p \) and \( r \) to maximize respective Stage 2 profits. Solving the system of equations given by the first-order conditions \( \frac{\partial \Pi^{AP}(x^{AP}, p, q, t)}{\partial p} = 0 \) and \( \frac{\partial \Gamma^{AP}(x^{AP}, r, \varrho)}{\partial r} = 0 \), we obtain the following best responses to the quality and technology investments in the preceding stages:

\[
\begin{align*}
r^{AP}(q, t, \varrho) &= \begin{cases} 
\frac{-q - d\sqrt{t} \delta + 2d\delta\sqrt{\varrho}}{3} & \text{if } q < \bar{q} := -\delta d(\sqrt{t} - 2\sqrt{\varrho}), \\
0 & \text{if } q \geq \bar{q}.
\end{cases} \\
p^{AP}(q, t, \varrho) &= \begin{cases} 
\frac{q - 2d\sqrt{t} \delta + d\delta\varrho}{3} & \text{if } q < \bar{q}, \\
\frac{-\delta d\sqrt{t}}{2} & \text{if } q \geq \bar{q}.
\end{cases}
\end{align*}
\]

As in the base model, the PDM optimally chooses its reward either according to its unconstrained best response or, if the unconstrained best response would yield a negative reward, according to the corner solution \( r^{AP} = \epsilon \), which is the minimum reward to ensure participation of consumers.

In Stage 1, anticipating prices and consumers’ usage, the CSP maximizes profit by either choosing a quality level to accommodate to the existence of the PDM or by choosing a monopoly-like quality, which induces the PDM to set a minimum reward of \( \epsilon \). Thus, the CSP’s optimal quality choice in Stage 1, given the investment levels \( t \) and \( \varrho \), is

\[
q^{AP}(t, \varrho) = \begin{cases} 
\frac{2}{3} \delta d(\sqrt{t} + \sqrt{\varrho}) & \text{if } t < \bar{t} := \frac{2}{\varrho} (2\sqrt{T\rho} + 9\rho), \\
\delta d\sqrt{t} & \text{if } t \geq \bar{t}.
\end{cases}
\]  

(4)

In Stage 0, the CSP and PDM simultaneously choose investments in data analytics technology, maximizing their respective profit and anticipating subsequent quality, price and usage levels. Solving the system of equations given by first-order conditions \( \frac{\partial \Pi^{AP}(x^{AP}, p^{AP}, q^{AP}, \varrho)}{\partial t} = 0 \)
and \( \frac{\partial \Gamma^{AP}(x^{AP},r^{AP},\varrho)}{\partial \varrho} = 0 \) yields the following profit-maximizing investment levels:

\[
\varrho^{AP} = \arg \max_{\varrho \in \mathbb{R}^{+}} \left[ \Gamma^{AP} \left( x^{AP}, r^{AP}, \varrho(t^{AP}) \right) \right] = \begin{cases} 
\frac{3d^{2}\delta^{2} \left( 3 \sqrt[3]{\bar{\alpha}} + \sqrt[3]{\frac{2}{21}} \sqrt[3]{\bar{\gamma}} \right)}{98 \sqrt{ab}} & \text{if } a > \bar{\alpha}, \\
\frac{d^{2}\delta^{2}}{4 \sqrt{ab} \sqrt{a}} & \text{if } a \leq \bar{\alpha},
\end{cases}
\]

\[
t^{AP} = \arg \max_{t \in \mathbb{R}^{+}} \left[ \Pi^{AP} \left( x^{AP}, p^{AP}, q^{AP}, t(\varrho^{AP}) \right) \right] = \begin{cases} 
\frac{d^{2}\delta^{2} \left( 21^{2/3} \sqrt[3]{\bar{\gamma}} + 7 \right)}{98a} & \text{if } a > \bar{\alpha}, \\
\frac{d^{2}\delta^{2}}{4a} & \text{if } a \leq \bar{\alpha}.
\end{cases}
\]

For \( a \leq \bar{\alpha} \), it is profitable for the CSP to choose a monopoly-like investment level \( t^{AP}_{\epsilon} \) (adjusted by the competitive impact of the PDM) and for the PDM to choose an investment level \( q^{MP,A,N}_{\epsilon} \), anticipating that it will remunerate consumers only with a minimum reward \( r^{AP} = \epsilon \). For \( a > \bar{\alpha} \), it is profitable for the PDM and the CSP to choose investment levels anticipating that the PDM will remunerate consumers with a reward \( r^{AP} > \epsilon \) in Stage 2.

Below, we show that no firm has an incentive to deviate from these strategy profiles given relative costs of the PDM and the CSP defined by \( \bar{\alpha}(b) \) and \( \bar{\alpha}(b) \).

In consequence, we obtain the following subgame-perfect equilibria:

**MINIMUM-INCOME EQUILIBRIUM** (denoted for \( \epsilon \to 0 \)):

\[
\varrho^{AP}_{\epsilon} = \frac{d^{2}\delta^{2}}{4a^{1/3}b^{2/3}}, \quad t^{AP}_{\epsilon} = \frac{d^{2}\delta^{2}}{4a}, \quad q^{AP}_{\epsilon} = \frac{d^{2}\delta^{2}}{2\sqrt{a}}, \quad p^{AP}_{\epsilon} = 0, \quad r^{AP}_{\epsilon} = 0,
\]

\[
x^{AP}_{\epsilon} = \frac{d^{2}\delta^{2}}{2\sqrt{a}}, \quad \Pi^{AP}_{\epsilon} = \frac{d^{4}\delta^{4}}{16a}, \quad \Gamma^{AP}_{\epsilon} = \frac{3d^{4}\delta^{4}}{16\sqrt{a^{2}b}}, \quad CS^{AP}_{\epsilon} = \frac{d^{4}\delta^{4}}{8a}.
\]
In the following, we first calculate feasibility conditions for the equilibrium candidates and then derive $	ilde{a}(b)$ and $\pi(b)$ by calculating the thresholds for which neither the CSP nor the PDM has an incentive to deviate from the respective equilibrium candidate by playing an out-of-equilibrium strategy. The minimum-income equilibrium candidate is only feasible if in Stage 2, given quality and technology investments, the PDM would choose a reward equal or below $\epsilon$ according to its unconstrained best response, thus opting for the corner solution. This is the case for for $a \leq b$, otherwise the PDM’s reward in Stage 2 exceeds $\epsilon$. On the contrary, the positive-income equilibrium candidate is only feasible if in Stage 2, given quality and technology investments, the PDM chooses a reward above $\epsilon$ ac-
cording to its unconstrained best response. This is the case for \( a > \frac{21}{64} b \), otherwise the PDM chooses the corner solution, violating expectations in earlier stages.

Next, we check whether the CSP or the PDM have an incentive and are able to deviate from the minimum-income equilibrium within the feasibility range determined by the condition \( a \leq b \). Therefore, we first test first whether the PDM can increase its profit \( \Gamma^A_P \) by choosing a \( q^{Dev} \) that maximizes \( \Gamma(x, r, p, q, t^A_P, q^{Dev}) \). This is only possible if \( q^{Dev} \) induces a positive \( r \) that exceeds \( \epsilon \) in Stage 2, because otherwise the equilibrium choice \( q^A_P \) is profit-maximizing by construction. Solving the necessary condition \( r(p, q, t^A_P, q^{Dev}) > 0 \) yields \( a > 1.5b \), which is outside of the feasibility range. Hence, the PDM is not able to profitably deviate from the minimum-income equilibrium within the feasible parameter range. We then test whether the CSP has an incentive to deviate from the minimum-income equilibrium. To do so, we check whether the CSP can increase its profit \( \Pi^A_P \) by choosing a pair \((q^{Dev}, t^{Dev})\) of Stage 0 and Stage 1 actions that maximizes \( \Pi(x, r, p, q^{Dev}, t^{Dev}, q^A_P) \). Solving the conditions \( \Pi(x, r, p, q^{Dev}, t^{Dev}, q^A_P) > \Pi^A_P \) and \( r(p, q^{Dev}, t^{Dev}, q^A_P) > 0 \) simultaneously yields the solution \( a > \tilde{a}(b) \), with the threshold \( \tilde{a}(b) \) being within the feasibility range of the minimum-income equilibrium (i.e., \( \tilde{a}(b) < b \)). Based on the derived conditions, we can conclude that the minimum-income equilibrium exists for \( a < \tilde{a}(b) \), because within this range no firm can profitably deviate from the equilibrium strategy configuration.

We then proceed analogously to check whether the CSP or the PDM have an incentive and are able to deviate from the positive-income equilibrium within the feasibility range given by \( a > \frac{21}{64} b \). First, for the PDM it can be shown that it cannot profitably deviate from the positive-income equilibrium, i.e., there is no \( q^{Dev} \) that yields a feasible solution, which simultaneously satisfies the necessary conditions \( \Gamma(x, r, p, q, t^A_P, q^{Dev}) > \Gamma^A_P \) and \( r(p, q, t^A_P, q^{Dev}) < 0 \). In contrast, we show that the CSP can profitably deviate from the positive-income equilibrium within the feasible range. More precisely, we show that the CSP can choose a pair \((q^{Dev}, t^{Dev})\) of Stage 0 and Stage 1 actions, such that the conditions \( \Pi(x, r, p, q^{Dev}, t^{Dev}, q^A_P) > \Pi^A_P \) and \( r(p, q^{Dev}, t^{Dev}, q^A_P) < 0 \) are satisfied simultaneously for the set of solutions defined by \( a < \tilde{a}(b) \approx 0.807b \). As \( \tilde{a}(b) > \frac{21}{64} b \), the threshold \( \tilde{a}(b) \) is within the feasible range of the positive-income equilibrium. Based on the derived conditions, we can
conclude that the positive-income equilibrium exists for $a > \tilde{a}(b)$, because within this range no firm can profitably deviate from the equilibrium strategy configuration.

**CSP’s quality with PDM:** The CSP’s quality in the minimum-income equilibrium never exceeds the monopoly quality. To show this, we solve $q^A = q_{AP}^\epsilon$. It is easy to see that the condition is only solved for the trivial solutions $d = 0$ and $\delta = 1$ and the unfeasible solution $\delta = -1$. Otherwise, within the feasible parameter range, $q^A > q_{AP}^\epsilon$. In contrast, quality in the positive-income equilibrium may exceed the monopoly quality if the cost advantage of the PDM is sufficiently large. To show this, we solve $q^A < q_{AP}^\epsilon$. Then, any solution must satisfy the necessary condition $a > \frac{875}{72} b$. In addition, for the solution to exist within the feasible parameter range $\delta$ must be sufficiently high.

**CSP’s price with PDM:** In the minimum-income equilibrium the CSP’s price is always set to zero (for $\epsilon \rightarrow 0$), which is identical to the price in the monopoly scenario. In the positive-income equilibrium, the CSP’s price is positive, and thus larger than the price in the scenario without PDM, for $a > \frac{448}{243} b$, which is the unique feasible solution that satisfies the condition $p^A < p_{AP}^\epsilon$.

**CSP’s profit with PDM:** The CSP’s profit in the minimum-income equilibrium never exceeds its profit in the monopoly scenario. To show this, we solve $\pi^A = \pi_{AP}^\epsilon$. It is easy to see that the condition is only solved for the trivial solutions $d = 0$ and $\delta = 1$ and the unfeasible solution $\delta = -1$. Otherwise, within the feasible parameter range, $\pi^A > \pi_{AP}^\epsilon$. In contrast, the CSP’s profit in the positive-income equilibrium may exceed the monopoly profit if the cost advantage of the PDM is sufficiently large. To show this, we solve $\pi^A < \pi_{AP}^\epsilon$. Then, any solution must satisfy the necessary condition $a > \tilde{a}(b) \approx 0.807 b$, which coincides with the feasibility threshold for the positive-income equilibrium. In addition, for the solution to exist within the feasible parameter range $\delta$ must be sufficiently high.

**Consumer surplus with PDM:** Consumer surplus in the minimum-income equilibrium never exceeds consumer surplus in the monopoly scenario. To show this, we solve $CS^A = $


It is easy to see that the condition is only solved for the trivial solutions $d = 0$ and $\delta = 1$ and the unfeasible solution $\delta = -1$. Otherwise, within the feasible parameter range, $CS^A > CS^AP$.

In contrast, consumer surplus in the positive-income equilibrium may exceed consumer surplus in the monopoly scenario if the cost advantage of the PDM is sufficiently large. To show this, we solve $CS^A < CS^AP$. Then, any solution must satisfy the necessary condition $a > \frac{7}{162} \left( -147\sqrt[3]{3^{2/3}b} + 63\cdot 2^{2/3}\sqrt[3]{3b} + 325b \right) \approx 3.629b$. That is, for consumers to be possibly better off with a PDM, the PDM must have sufficiently lower investments costs for data analytics technology than the CSP.

**Symmetric cost structures:** For symmetric costs of the CSP and the PDM, i.e., $b = a$, consumer surplus with the PDM is always lower than without the PDM. To show this, we solve $CS^AP = CS^A$ assuming $b := a$ and show that there is no solution within the feasible parameter range, except for the trivial solution $d = 0$. It is then easy to verify that $CS^AP > CS^A$ for a specific set of values within the feasible parameter range.

With symmetric costs $b := a$, only the positive-income equilibrium of the two identified equilibrium types is a subgame-perfect equilibrium (as $a > \bar{a}(b)$ and $a > \overline{a})$. Thus, consumer surplus with and without the PDM is given by

$$CS^A = \frac{d^4}{8b} ,$$

$$CS^AP = \frac{9 \left( 16 + 9\sqrt[3]{2} + 3\cdot 21^{2/3} \right) d^4 \delta^4}{9604a} .$$

Solving $CS^AP = CS^A$ yields solutions $d = 0$, $\delta = -1.203$ and $\delta = 1.203$, with the latter two being outside of the feasible parameter range ($\delta \in [0, 1]$).