ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Gurkan, Huseyin; de Véricourt, Francis

**Working Paper**

# Contracting, pricing, and data collection under the AI flywheel effect

ESMT Working Paper, No. 20-01 (R2)

**Provided in Cooperation with:**
ESMT European School of Management and Technology, Berlin

*Suggested Citation:* Gurkan, Huseyin; de Véricourt, Francis (2021) : Contracting, pricing, and data collection under the AI flywheel effect, ESMT Working Paper, No. 20-01 (R2), European School of Management and Technology (ESMT), Berlin, https://nbn-resolving.de/urn:nbn:de:101:1-202105181406056638217 2

This Version is available at:
https://hdl.handle.net/10419/235147

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ISSN 1866-3494

May 17, 2021

**ESMT Working Paper 20-01 (R2)**

# Contracting, pricing, and data collection under the AI flywheel effect

Huseyin Gurkan, ESMT Berlin

Francis de Véricourt, ESMT Berlin

Revised version

Copyright 2021 by ESMT European School of Management and Technology GmbH, Berlin, Germany, https://esmt.berlin/.

All rights reserved. This document may be distributed for free – electronically or in print – in the same formats as it is available on the website of the ESMT (https://esmt.berlin/) for non-commercial purposes. It is not allowed to produce any derivates of it without the written permission of ESMT.

Find more ESMT working papers at ESMT faculty publications, SSRN, RePEc, and EconStor.

# Contracting, Pricing and Data Collection under the AI Flywheel Effect

Huseyin Gurkan
ESMT Berlin, Germany, huseyin.gurkan@esmt.org,

Francis de Véricourt
ESMT Berlin, Germany, francis.devericourt@esmt.org,

This paper explores how firms that lack expertise in machine learning (ML) can leverage the so-called AI Flywheel effect. This effect designates a virtuous cycle by which, as an ML product is adopted and new user data are fed back to the algorithm, the product improves, enabling further adoptions. However, managing this feedback loop is difficult, especially when the algorithm is contracted out. Indeed, the additional data that the AI Flywheel effect generates may change the provider's incentives to improve the algorithm over time. We formalize this problem in a simple two-period moral hazard framework that captures the main dynamics among ML, data acquisition, pricing, and contracting. We find that the firm's decisions crucially depend on how the amount of data on which the machine is trained interacts with the provider's effort. If this effort has a more (less) significant impact on accuracy for larger volumes of data, the firm underprices (overprices) the product. Interestingly, these distortions sometimes improve social welfare, which accounts for the customer surplus and profits of both the firm and provider. Further, the interaction between incentive issues and the positive externalities of the AI Flywheel effect has important implications for the firm's data collection strategy. In particular, the firm can boost its profit by increasing the product's capacity to acquire usage data only up to a certain level. If the product collects too much data per user, the firm's profit may actually decrease, i.e., more data is not necessarily better. As a result, the firm should consider reducing its product's data acquisition capacity when its initial dataset to train the algorithm is large enough.

*Key words*: Data, Machine Learning, Data Product, Pricing, Incentives, Contracting

## 1 Introduction

To train ML algorithms, companies often deploy their artificial intelligence (AI)-based products early and collect usage data from their first customers. As new data are fed back to the algorithm, the technology improves, enabling further adoptions and thus the acquisition of additional data. This virtuous feedback loop, sometimes referred to as the AI Flywheel effect in the popular press

(Trautman 2018), compounds the economic effect by which quality increases demand, according to the statistics principle whereby data improves accuracy.

The AI Flywheel effect has many applications, from voice recognition systems (Sarikaya 2019) to self-driving vehicles (Miller 2016), and even explains how certain web search engines ended up dominating the market. Nonetheless, this virtuous cycle is perhaps most useful for smaller teams or novel and specialized applications for which data are scarce. The founders of the startup Blue River Technology famously established their first dataset manually to train an AI system that would distinguish weeds from crops, a crucial step to optimize pesticide sprays in farming (Ng 2018, Trautman 2018). This yielded an algorithm with low performance, but with its adoption by early users, the company could leverage the effects and significantly improve the algorithm. The company was sold in 2017 for more than $300 million (Golden 2017).

However, despite its apparent simplicity, the AI Flywheel effect is difficult to implement, especially among the small organizations that would most benefit from it. First, a firm makes choices and, in particular, pricing decisions that affect demand alongside accuracy and hence interfere with the virtuous cycle. More specifically, the AI Flywheel effect introduces an additional tradeoff between improving the accuracy of algorithms and maximizing revenue, which the firm needs to consider when setting its pricing strategy.

Second, and perhaps more importantly, many firms lack the expertise to develop ML algorithms. Indeed, the economy has experienced a significant shortage of skilled data scientists, which particularly affects startups and small organizations (Nicolaus Henke et al. 2016). This shortage has given rise to a striving outsourcing industry (Research Nester Pvt. Ltd 2019), and many startups have achieved success by outsourcing their technology (examples include Skype, Opera, and Slack, to name a few; see Cengiz 2015).

However, relying on outsourcing gives rise to incentive issues, which may impair accuracy and thus again interfere with the AI Flywheel effect. For example, a provider may shirk by applying standard third-party software that may be suboptimal for the task or may expose the firm to threats (Kendra et al. 2019, Bursztein 2018). If these algorithms are not developed with care, they only learn surface statistical regularities, which affects their ability to generalize and thus their accuracy (Jo and Bengio 2017). More generally, the provider may have an incentive to wait for more usage data before exerting any effort to improve the algorithm. In fact, incentive issues such as these may not disappear if the firm does not outsource the algorithm but instead employs an expert. Indeed, the details of ML algorithms and their outputs notoriously suffer from a lack of explainability (Lipton 2016, Ribeiro et al. 2016), rendering the expert's efforts to improve accuracy difficult to observe and contract on.

Third, the amount of data to which the provider has access for training the algorithm may exacerbate the incentive issues. For instance, research in AI has suggested that accuracy depends less on the specifics of the algorithm and more on the data on which it is trained as data volume increases (Banko and Brill 2001, Halevy et al. 2009). In this case, the provider's effort to develop the machine matters more when data are scarce, that is, in the early stages of the AI Flywheel effect. Hence, the intensity of the incentive issues may change over time as the feedback loop between accuracy and usage data unfolds.

The goal of this paper is to shed light on how firms that lack expertise in ML can leverage and optimize the AI Flywheel effect. In this context, we seek to understand how the need to mitigate the incentive issues created by outsourcing the algorithm affect the firm's pricing and data collection strategies. To that end, we formalize the problem in a simple two-period moral hazard framework, which captures the previous three features: the accuracy vs. revenue tradeoff, the incentive issues, and the impact of data on the intensity of these issues. In particular, we draw on the ML literature to represent how the accuracy of an algorithm depends on the amount of data on which it is trained.

Our analysis reveals that the firm's decisions crucially depend on how the amount of data on which the machine is trained interacts with the provider's effort. Specifically, if the provider's effort has a more significant impact on accuracy for larger volumes of data, the firm *underprices* the product in order to acquire more data from the market, retrain the algorithm, and generate more revenues in the future. By contrast, the firm *overprices* and collects less data if the provider's effort is most impactful when data are scarce. These effects also affect social welfare and customer surplus, which improve (deteriorate) when the impact of the provider's effort on accuracy increases (decreases) with more data.

More importantly, the interaction between the positive externalities of the AI Flywheel effect and the incentive issues can alter the firm's data collection strategy. In particular, a key and specific design choice for data products is the amount of data that the product can collect on its user, which we refer to as the data acquisition capacity in the following (Spencer 1990, Kos et al. 2012). The firm can increase this amount by relying on third-party services in the case of mobile and web applications (Deshpande 2019) or increasing the capacity of embedded sensors in the case of physical products (McGrath and Scanaill 2013). This, in turn, should provide the firm with more data overall to improve its algorithm and ultimately increase profits.

We find, however, that increasing the product's data acquisition capacity sometimes reduces the total amount of data collected by the firm. This happens when the provider's effort has a less significant impact on accuracy for larger volumes of data. In this case, increasing the data acquisition capacity does boost profit, but up to a certain point. If the product collects too much

data per user, the profit may actually decrease. In other words, more data is not necessarily better. This further implies that the firm should consider reducing the product's data acquisition capacity once its initial dataset to train an algorithm is large enough.

These findings are overall robust to a variety of settings. We show that they hold in a competitive provider market or when the firm let the provider shirk in the beginning of the algorithm's training process (i.e., in the first period of our setup). The firm may also be able to withhold a portion of the collected data from the provider. In this case, the firm needs to decide how much data to share, in addition to its data collection and pricing strategies. In this setting, we show that improving the acquisition capacity too much does not bring any additional benefit when the data impact is decreasing.

Taken together, these results characterize how the interaction between the incentive issues and the positive adoption externalities of the AI Flywheel effect significantly affects the firm's pricing and data collection strategies. One key driver of these effects is the impact of data on the provider's effort when training the algorithm.

After reviewing the literature in Section 2 and presenting the model in Section 3, in Section 4 we characterize the deviations from first best arising from incentive issues . We then analyze the impact of improving the product's data acquisition capacity on the firm's data collection strategy and its profit in Section 5. We explore the robustness of our findings in Section 6 and discuss their managerial implications along with future research directions in Section 7.

## 2 Literature Review

The advent of the digital economy has recently generated new research on data privacy and markets in both management science and economics. This new stream of research explores the impact of data leakage in platform business models (Acemoglu et al. 2019), the issue of selling data (Bimpikis et al. 2019, Mehta et al. 2019), and the effect of collecting data on privacy and price discrimination (Loertscher and Marx 2019). By contrast, our work focuses on the outsourcing of ML algorithms that make use of this data, which creates incentive issues that dynamically interact with the amount of available data. More generally, our study is the first in this stream of research to explore the problem of managing the AI Flywheel effect.

Our work is also related to the large literature on dynamic pricing with learning. Of particular interest is the recent stream of research on new experience goods and quality learning. Yu et al. (2015), for instance, study the dynamic pricing of new experience goods in the presence of two-sided learning (learning about quality via consumer reviews). In their setting, the pricing decision affects both revenue and the flow of information. They show that consumer-generated quality information may decrease the firm's profit and even consumer surplus. Feldman et al. (2018) also analyze the

pricing and quality design of new experience goods for consumers who are social learners. They characterize the deviation of the firm's optimal policy from a setting without social learning. In these setups, quality is a decision variable that is set ex ante, and learning concerns either the firm learning about the consumers or the consumers learning about quality. By contrast, accuracy is dynamically improved in our setting and learning concerns the training of the algorithm, that is, the enhancement of quality. In addition, we consider moral hazard issues, which is not the focus of this stream of research.

Another related stream of the literature studies the positive network effects (or positive adoption externalities) that customers derive from the consumption of the same product by others (see Katz and Shapiro 1994 and Shy 2011 for a general review). In particular, this literature considers the dynamic monopoly pricing of durable goods with network effects that are close to our first-best benchmark (see for instance Cabral et al. 1999, Bensaid and Lesne 1996, Gabszewicz and Garcia 2008). In these studies, customers are strategic and long-lived and thus have incentives to postpone their purchases in order to benefit from the network effects generated by others' consumption. This induces the firm to decrease the price in the first period.

Our work differs from this stream of research for two reasons. First, we explore whether a firm should strengthen the network effects, which corresponds to improving the data acquisition capacity in our setup. Although the acquisition capacity is a natural design choice in our setting, firms cannot easily affect the intensity of the network effects in the aforementioned literature. Second, these studies are not concerned with incentive issues on the provider side. In our setup, however, the amount of collected data interacts with both the positive adoption externalities of the AI Flywheel effect and the incentive issues that developing the product brings about.

From a more technical point of view, our model is a dynamic moral hazard problem with binary effort choices and binary outcomes. Different versions of this problem have been studied, especially in the sales force management literature. Schöttner (2016) analyzes a multi-period setting with different sales probabilities in each period when the firm can obtain only aggregate information on sales. Despite being different across periods, the sales probabilities are taken as constant. Kräkel and Schöttner (2016) consider a two-period model with binary effort choices and analyze the optimal contracts. They also study a case where the second-period sales opportunity randomly depends on the outcome of the first period with exogenous probabilities. Schmitz (2005, 2013) explore similar settings, where the probabilities of success are assumed to change across periods.

In addition, although these papers explore various configurations of the uncertainty structure, the probability of observing a favorable outcome in their models is fixed (Schöttner 2016) or exogenously depends on the outcome in the first period (Kräkel and Schöttner 2016, Schmitz 2005, 2013). By contrast, these probabilities are fully endogenous in our setup.

In this stream of research, Dai and Jerath (2013, 2019) study a slightly more general dynamic moral hazard problem with binary effort choices and three possible outcomes. A key aspect of this model, which is also discussed in de Véricourt and Gromb (2018, 2019) for more general distributions of outcomes, is that the firm's capacity decision interacts with the moral hazard problem. Indeed, in Dai and Jerath (2013), de Véricourt and Gromb (2018), Dai and Jerath (2019), and de Véricourt and Gromb (2019), a low capacity level may censor high demand realizations, which exacerbates the incentive issue. More generally, deviations from the first-best setting are observed in other contexts as well, (see, e.g., Alles et al. 1995 for over/understocking decisions in just-in-time production systems). Indeed, a common theme of the moral hazard literature is that it is attractive for the principal to take actions that mitigate incentive issues and the unobservability of the agent's effort in particular.

Our work considers a different type of interaction between the firm's decision (pricing in our case) and the moral hazard problem through a monotonic property of the data impact ratio. Moreover, ours is the first to articulate this theme in the context of data analytics. Specifically, the AI Flywheel effect endogenously interacts with incentive issues, and both are affected by data. Thus, the firm's ability to leverage the AI Flywheel effect is constrained by the incentive misalignment between the firm and the provider.

Our work also contributes to the rich operations management literature in entrepreneurship, as our setup is particularly relevant for cash-constrained firms with a lack of technical skills. The points of focus in this literature widely range from investment timing (Swinney et al. 2011) and financial capabilities (Tanrısever et al. 2012) to complementary technologies (Anderson Jr and Parker 2013). By contrast, we provide insight on how cash-constrained firms can leverage a business model based on the AI Flywheel effect.

Finally, a key feature of our problem is the effect of the amount of data collected by the firm on the algorithm's accuracy. The ML literature often represents this link between data and accuracy with a *learning curve* (see, e.g. Perlich et al. 2003, Gu et al. 2001, Leite and Brazdil 2005, Figueroa et al. 2012). These curves are commonly estimated with increasing concave exponential or power functions (see, for instance, Table 1 in Viering and Loog 2021 for a list of references). In our setup, we draw on this literature to represent a learning curve as a (non-parametric) increasing concave mapping.

## 3 Model Description

We model the problem of managing the AI Flywheel effect in an elementary two-period moral hazard framework. In our setup, the firm (the principal) outsources the development and training of the algorithm to a provider (the agent) at the beginning of each period. The resulting accuracy

of the algorithm depends on both the provider's effort and the size of the available dataset. Given this accuracy level, the firm markets the product to users with heterogeneous accuracy and price sensitivities. At the end of each period, demand is realized, which determines the profit and the additional generated data for that period.

### 3.1 Data, Accuracy, and Revenue

At the beginning of period $t = 1, 2$, the size of the available dataset to train the algorithm is equal to $d_{t-1}$. In particular, $d_0$ denotes the size of the firm's initial dataset. In each period $t$, the provider can either exert effort $e = w$ at cost $\kappa > 0$ or shirk $e = s$ at no cost. We denote by $\alpha_t$ the resulting algorithm's accuracy, which can be either high or low with $\alpha_t \in \{\alpha_h, \alpha_\ell\}$, where $\alpha_h > \alpha_\ell$.[1]

Given the data size $d_{t-1}$ and effort $e$, the probability of high accuracy ($\alpha_t = \alpha_h$) is equal to $\pi_e(d_{t-1})$ with $\pi_w(d_{t-1}) > \pi_s(d_{t-1})$.[2] Consistent with the ML literature (see Section 2), we refer to $\pi_e : [d_{\min}, d_{\max}] \to [0, 1)$ for $e \in \{s, w\}$ as *learning curves* and assume that they are twice-differentiable, increasing concave functions with a continuous second-order derivative. In other words, the accuracy of ML models increases with data, but the marginal effect of additional data is decreasing (see Banko and Brill 2001, for instance). The further lower bound $d_{\min} > 0$ is the minimum size required for the development of a functioning ML model, and $d_{\max} < \infty$ is the largest possible total data size on which the algorithm can be trained.

Given accuracy $\alpha_t$, the firm then prices and markets the product. The market in period $t$ corresponds to a continuum of buyers of total mass normalized to one, a common framework in the pricing literature (e.g., Aflaki et al. 2019, Feldman et al. 2018, Yu et al. 2015). Each buyer has a private accuracy sensitivity $v$ that is drawn from the standard uniform distribution with c.d.f. $F$, p.d.f. $f$, support $[0, 1]$, and a virtual value function $\phi(v) = v - \bar{F}(v)/f(v)$, where $\bar{F}(\cdot) = 1 - F(\cdot)$. Hereafter, we use the notation $\bar{x}$ to denote $1 - x$ for an arbitrary term $x$. A buyer with sensitivity $v$ purchases the product with accuracy $\alpha$ and price $p$ if $\alpha v - p \geq 0$. Given the accuracy-price pair $(p, \alpha)$, demand is equal to $\bar{F}(p/\alpha)$, which yields revenue $p\bar{F}(p/\alpha)$.[3] Equivalently, the firm may

[1] Accuracy in our setup designates the overall accuracy, which includes both "Sensitivity" and "Specificity", i.e., *Accuracy* = *Sensitivity* × # of positives/(# of positives + # of negatives) + *Specificity* × # of negatives/(# of positives + # of negatives) (see Tharwat 2020, Sokolova et al. 2006). This metric is typically used to measure the performance of ML algorithms (see, e.g., Ferri et al. 2009, Blagec et al. 2020).

[2] We choose a non-parametric representation of these learning curves because our focus is on the effect of data collection on accuracy and we want to tease out which fundamental properties of these functions drive our results. This, however, implies that closed-form solutions cannot be obtained and that our results are all derived from implicit functions.

[3] Note that buyers can learn the accuracy via word of mouth and other network effects as is typical for digital products, which gives rise to a diffusion process in the market (see, e.g., Campbell 2013). To keep the model tractable, we have abstracted away from these dynamics and assume instead that this process is fast enough so that most buyers learn the accuracy before purchase.

choose demand quantity $q$ instead of price $p$, with $p = \alpha F^{-1}(\bar{q})$ and, hence, $p\bar{F}(p/\alpha) = q\alpha F^{-1}(\bar{q})$, where $F^{-1}$ is the inverse of the c.d.f. $F$.

A key attribute of data-enhanced products is the amount of data that the product can collect on its user. The firm may increase this amount by relying on third-party services in the case of mobile and web applications (Deshpande 2019) or increasing the capacity of embedded sensors in the case of physical products (McGrath and Scanaill 2013). We refer to this quantity as the data acquisition capacity $\nu > 0$ in the following, which corresponds to the expected amount of data the firm collects per user.

Given demand $q$, therefore, the total amount of collected data $\delta$ in the period is equal to $\delta = \nu q$. Thus, at the end of period $t = 1, 2$ the size of the available dataset is equal to $d_t = \delta_t + d_{t-1}$. Further, the firm's revenue can be expressed in terms of acquisition capacity $\nu$, collected data $\delta$, and accuracy $\alpha$. Specifically, we define this revenue as $R_\nu(\delta, \alpha)$ with $R_\nu(\delta, \alpha) \triangleq \alpha\delta/\nu F^{-1}((1 - \delta/\nu))$.

Thus, decreasing price increases demand and hence the amount of collected data but may also decrease revenue $R_\nu$. In this sense, our model captures the tradeoff associated with the AI Flywheel effect between maximizing revenue and collecting additional data. In addition, parameters $(d_0, \nu)$ characterize the potential strength of the AI Flywheel effect in our setup. Indeed, we have $d_1 = \nu q + d_0$; hence, data size $d_0$ specifies the firm's starting point in the virtuous cycle, while acquisition capacity $\nu$ influences the speed at which the firm can leverage this cycle.

Figure 1 depicts the timing of the events corresponding to our setup.

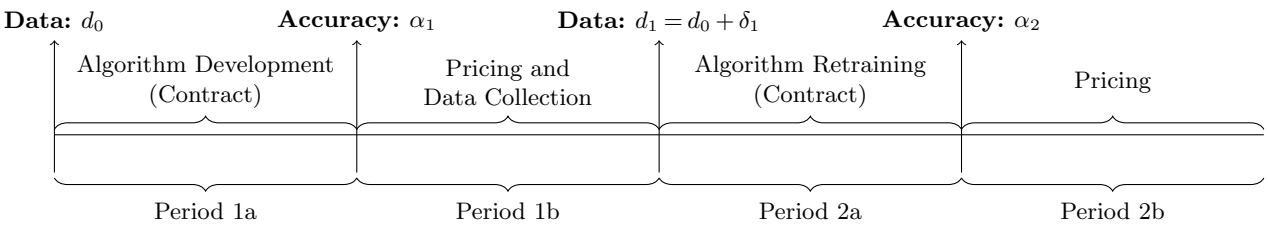
*Period 1a – Initial Algorithm Development.* The firm starts with an initial dataset of size $d_0$. Based on this dataset, the provider chooses effort $e \in \{s, w\}$ to develop a first version of an ML algorithm. The algorithm's accuracy $\alpha_1$ is then realized according to probability $\pi_e(d_0)$.

*Period 1b – Pricing and Data Collection.* Given accuracy $\alpha_1$, the firm prices and markets the product. The firm collects $\delta_1$ such that the total size of the dataset becomes $d_1 = d_0 + \delta_1$ and generates revenue $R_\nu(\delta_1, \alpha_1)$.

*Period 2a – Algorithm retraining and improvement.* If $\alpha_1 = \alpha_\ell$, the provider retrains the algorithm with an augmented dataset of size $d_1$. (Otherwise, the maximum possible accuracy level $\alpha_h$ is achieved, and the firm does not need the provider to improve accuracy further.)[4] The provider again chooses effort $e$ to retrain the algorithm, which yields accuracy $\alpha_2$ according to probability $\pi_e(d_1)$. The probability of achieving high accuracy $\alpha_h$ increases in this period, that is, $\pi_e(d_1) > \pi_e(d_0)$, because of the dataset increase $d_1 \geq d_0$, and since more data improves accuracy, i.e., $\pi_e(\cdot)$ for $e \in \{s, w\}$ are increasing functions.

[4] Our model and results easily extend to the case where accuracy is cumulative, that is, in situations where accuracy can be further improved when $\alpha_1 = \alpha_h$.

**Figure 1** **Sequence of events**




*Period 2b – Pricing.* Finally, given accuracy $\alpha_2$, the firm prices and markets the product, which determines $\delta_2$ and generates revenue $R_\nu(\delta_2, \alpha_2)$.

At the end of the time horizon, the firm has no further incentive to retrain the algorithm using additional data $\delta_2$, which also means that $d_1$ corresponds to the size of the largest dataset on which the algorithm is ultimately trained. The maximum possible size of this dataset is then equal to $d_{\max} = \nu_{\max} + d_0$ (recall that the market size is normalized to one), where $\nu_{\max}$ is the maximum acquisition capacity.

Taken together, our setup captures the AI Flywheel effect: more data at the end of Period 1, $d_1$, leads to a higher probability $\pi_e(d_1)$ of achieving high accuracy $\alpha_h$ in Period 2 (for any effort $e \in \{s, w\}$), which in turn leads to more demand in Period 2 in expectation (given fixed $p_2$). The firm may foster or suppress this effect by adjusting its pricing strategy.

### 3.2 The Data Impact

Thus far, we have ignored the incentive issues that outsourcing the development of the algorithm creates. In particular, the accuracy level depends on the available data and the provider's effort. Thus, the availability of data may interact with the intensity of the moral hazard problem that the firm faces in each period. To characterize this interaction between effort and data, we introduce the notion of the *data impact*, which we denote by $\rho(d)$. The data impact maps dataset size $d$ to the normalized effect of shirking on the probability of high accuracy, that is,

$$\rho(d) \triangleq \frac{\pi_w(d) - \pi_s(d)}{\pi_w(d)}\,. \tag{1}$$

We further assume that $1/\rho(d)$ is convex in $d$, which essentially requires data impact $\rho(\cdot)$ not to be too convex. This technical restriction is milder than log-concavity and hence concavity.

Overall, when data impact $\rho(d)$ is constant in $d$, the effect of shirking on accuracy is independent of the data size on which the algorithm is trained. However, when the data impact increases (decreases) in $d$, exerting effort increases the probability of high accuracy more (less) with more data.

Many corroborative evidence supports the assumption that the data impact is decreasing in a variety of practical settings. Specifically, empirical evidence demonstrates that the difference in

accuracy between sophisticated ML algorithms (which require more effort to develop and train) and more crude ones decreases with the amount of data on which they are trained (Banko and Brill 2001; Halevy et al. 2009). Practitioners have also recognized this phenomenon; see, for example, Gutierrez (2016).[5]

Yet, certain circumstances may justify the assumption of increasing data impact. This is the case, for instance, when biases in the training data sets are of particular concern. The impact of these biases is typically more pronounced for larger datasets.[6] In this case, the effort to improve the algorithm has a bigger impact with more data. In addition, deep learning algorithms seem to outperform very basic statistical models (such as logistic regressions) to a greater magnitude when datasets are larger (see Hackathorn 2018, for instance). Thus, the data impact should increase when the firm does not have the skills to run nor check these more basic methods and hence cannot detect this form of shirking.

Finally, the monotonicity of the data impact is related to the monotone likelihood ratio property (MLRP), which is commonly assumed in the moral hazard literature. In our setup, the MLRP property corresponds to $\pi_w(d)/\pi_s(d) \geq \bar{\pi}_w(d)/\bar{\pi}_s(d)$ for a given $d$ and holds for all $d$ since effort always improves accuracy $\pi_w(d) > \pi_s(d)$ (see Dai and Jerath 2019, for instance). Loosely speaking, the property guarantees that high accuracy is more indicative of high effort *ex post*. The monotonicity of the data impact determines the magnitude of this effect *ex ante* for the amount of available data to train the algorithm.

### 3.3 Contracts

The firm faces a moral hazard problem in each period and has commitment power across periods. Thus, the firm needs to offer a single contract that determines the provider's payments in the first period along with the expected payoff in the second period through the total amount of collected data $\delta_1$ and the second period payments (we consider the noncommitment case in Section 6.2).

Accuracy realizations $\alpha_t$, $t = 1, 2$, are contractable, but effort is not. In particular, the product may track (ex-post) accuracy once it is distributed, or the firm can always hold some data to test the accuracy. Alternatively, the firm may contract on revenues, which is equivalent to contracting on accuracy in our setup.

By contrast, ML algorithms are notoriously difficult to explain to non-experts (Lipton 2016, Ribeiro et al. 2016), rendering the expert's efforts to improve accuracy difficult to observe and

[5] For instance, the following quote is from a data scientist at The Trade Desk: "The bigger the training set, the better the model. But less obviously, and more importantly, the difference between a fancy algorithm and a simple one decreases with more data." Gutierrez (2016).

[6] A well-known example is Microsoft's AI-based chatbot in Twitter, which resulted in unintended tweets as the size of the dataset feeding the algorithm increased, see Vincent (2016).

contract on (Gromb and Martimort 2007). Thus, the contract's payments in each period depend on the public history and are contingent upon the accuracy realizations.[7] Specifically, the history in Period 1a decreases to $d_0$, and the first contract consists of payments $x_{1\ell}(d_0)$ and $x_{1h}(d_0)$ that are made if $\alpha_1 = \alpha_\ell$ and $\alpha_1 = \alpha_h$, respectively. Given realization $\alpha_1$, the firm collects additional data, which yields size $d_1$ at the end of the period. The history in Period 2a is then $(d_1, \alpha_1)$, and the second contract consists of payments $x_{2\ell}(d_1, \alpha_1)$ and $x_{2h}(d_1, \alpha_1)$, which again correspond to high and low accuracy levels, respectively.[8] The provider is further protected by its limited liability; thus, $x_{1\ell}, x_{1h}, x_{2\ell}$ and $x_{2h}$ are all non-negative.

### 3.4 The Firm's Problem

The firm's problem is to maximize the total expected profit, which is the expected revenue net of payments over both periods, subject to incentive compatibility constraints.

We assume that the optimal price neither covers nor excludes the entire market (see Section EC.1 of the e-companion for formal conditions) and that the firm prefers the provider to exert effort in both periods, which holds if $\kappa \leq \bar{\kappa}$ and $\nu \leq \bar{\nu}$ for positive thresholds $\bar{\kappa}$ and $\bar{\nu}$ (see Section EC.3 of the e-companion for the existence of these thresholds).[9] These assumptions are made for the sake of simplicity; see, for instance, Laffont and Martimort (2009), and Feldman et al. (2018), Choudhary et al. (2005), respectively.

We formulate this problem via backward induction starting from the second period (see Figure 1). We then denote by $J_{2b}(\alpha_2)$ the firm's optimal expected profit in Period 2b given accuracy $\alpha_2$, such that

$$J_{2b}(\alpha_2) = \max_{\delta_2 \in [0,\nu]} R_\nu(\delta_2, \alpha_2)\,. \tag{2}$$

The firm chooses the amount of collected data (or equivalently the price) so as to maximize the expected revenue in the current period. As there is no continuation, we refer to this problem as the *myopic* problem. In particular, size $d_1$ does not play any role in this problem, which corresponds to situations where both the AI Flywheel effect and the moral hazard problem are absent. We denote by $\delta^{\mathtt{M}}$ the value of $\delta_2$ that solves Problem (2).[10]

[7] In our setup, the accuracy realization affects the demand per Section 3. Thus, the contract's payments can equivalently be contingent upon the demand.

[8] Recall that these last payments are only meaningful when $\alpha_1 = \alpha_\ell$, as no contract is required in the second period when $\alpha_1 = \alpha_h$.

[9] Threshold $\bar{\nu}$ can be infinite depending on the problem parameters; see Proposition EC.5 in Section EC.3 of the e-companion.

[10] Alternatively, $\delta_2$ can be carried to Problem (6) as a decision variable. However, its optimal value $\delta^{\mathtt{M}}$ remains the same because $\delta_2$ does not intervene with the other decision variables in Problem (6).

Similarly, we denote by $J_{2a}(\alpha_1, d_1)$ the firm's expected continuation profit in Period 2a, given payments $x_{2h}, x_{2\ell}$, and accuracy $\alpha_1$ and data size $d_1$, such that

$$J_{2a}(\alpha_\ell, d_1) = \pi_w(d_1)[J_{2b}(\alpha_h) - x_{2h}] + \bar{\pi}_w(d_1)[J_{2b}(\alpha_\ell) - x_{2\ell}] \tag{3}$$

$$J_{2a}(\alpha_h, d_1) = J_{2b}(\alpha_h). \tag{4}$$

When $\alpha_1 = \alpha_\ell$, the firm needs to set payments such that the provider has enough incentives to exert effort, as will be formalized by incentive constraint (9). These payments are then deduced from the firm's expected revenues in Period 2a. Here, data size $d_1$ affects the chance of improving accuracy in the next period via probabilities $\pi_e(\cdot)$, $e \in \{w, s\}$. When $\alpha_1 = \alpha_h$, recall that the firm does not need nor pay the provider.

Moving to the first period, we denote by $J_{1b}(\alpha_1, d_0, \delta_1)$ the firm's expected continuation profit in Period 1b given accuracy $\alpha_1$, data size $d_0$, and collected data $\delta_1$, such that

$$J_{1b}(\alpha_1, d_0, \delta_1) = R_\nu(\delta_1, \alpha_1) + J_{2a}\big(\alpha_1, d_0 + \delta_1\big) \text{ for } \alpha_1 \in \{\alpha_h, \alpha_\ell\}. \tag{5}$$

When $\alpha_1 = \alpha_\ell$, the choice of data $\delta_1$ (or equivalently the price) affects current revenues directly and future ones indirectly by increasing the dataset size to $d_0 + \delta_1$.

We are now ready to define the overall firm's problem. Given initial data size $d_0$, we denote by $J_{1a}(d_0)$ the optimal total expected profit in Period 1a, such that

$$J_{1a}(d_0) = \max_{\substack{x_{1h}, x_{1\ell} \geq 0 \\ x_{2h}, x_{2\ell} \geq 0 \\ \delta_{1h}, \delta_{1\ell} \in [0, \nu]}} \pi_w(d_0)[J_{1b}(\alpha_h, d_0, \delta_{1h}) - x_{1h}] + \bar{\pi}_w(d_0)[J_{1b}(\alpha_\ell, d_0, \delta_{1\ell}) - x_{1\ell}] \tag{6}$$

s.t.

$$\pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)\left[x_{1\ell} + J_p(d_0)\right] - \kappa \geq \pi_s(d_0)x_{1h} + \bar{\pi}_s(d_0)\left[x_{1\ell} + J_p(d_0)\right] \tag{7}$$

$$J_p(d_0) = \pi_w\left(d_0 + \delta_{1\ell}\right)x_{2h} + \bar{\pi}_w\left(d_0 + \delta_{1\ell}\right)x_{2\ell} - \kappa \tag{8}$$

$$J_p(d_0) \geq \pi_s(d_0 + \delta_{1\ell})x_{2h} + \bar{\pi}_s(d_0 + \delta_{1\ell})x_{2\ell}\,. \tag{9}$$

The firm needs to provide incentive to the provider for both periods. The expected payments that the contract of the second period brings about affect the provider's incentives in the first period. Specifically, $J_p(\cdot)$ in (7) corresponds to the provider's expected continuation profit, which is equal to the expected optimal payments in the second period net of the effort cost; see (8). A key aspect of our setup is that the distribution of these future payments, $\pi_w(d_1)$, depends on the choice of $\delta_{1\ell}$ since $d_1 = d_0 + \delta_{1\ell}$. Thus, the choice of $\delta_{1\ell}$ (or equivalently price) not only makes the tradeoff between present and future revenues as in (5) but also determines the intensity of the incentive issue.

We next show that the optimal data sizes solving Problem (6) and hence the corresponding optimal payments are unique.

PROPOSITION 1. *The optimal solution to Problem* (6) *is unique such that* $\delta_{1h} = \delta^{\mathtt{M}}$ *and* $\delta_{1\ell} = \delta^*$, *where* $\delta^*$ *is the unique optimal solution of*

$$\max_{\delta \in [0,\nu]} \bar{\pi}_w(d_0)\left[R_\nu(\delta, \alpha_\ell) + \pi_w(d_0+\delta)(\alpha_h - \alpha_\ell)\tau\right] - \frac{\kappa}{\rho(d_0+\delta)}\,.$$

(The proof of Proposition 1 alongside the proofs of selected results are provided in the appendix. All other proofs are in the e-companion.) If the first version of the algorithm is already highly accurate ($\alpha_1 = \alpha_h$), no further improvement is necessary, and the firm does not need to deviate from the myopic price. We refer to $\delta^*$ as the optimal solution $\delta_{1\ell}$ of Problem (6) and to $p^*$ as the price that yields data size $\delta^*$ (see Section 3.1).

This proposition further reveals how the main effects captured by our model endogenously determine data collection size $\delta^*$. Indeed, deviating $\delta$ away from $\delta^{\mathtt{M}}$ decreases the firm's current period revenue $R_\nu(\delta, \alpha_\ell)$. However, increasing $\delta$ always boosts the positive externalities of the AI Flywheel effect ($\pi_w(d_0+\delta)(\alpha_h - \alpha_\ell)\tau$). Moreover, additional data may exacerbate or mitigate the incentive issues and associated agency cost $\kappa/\rho(d_0+\delta)$ depending on the monotonicity of the data impact. The choice of $\delta^*$ essentially balances these different tradeoffs simultaneously.

# 4 Impact of Incentive Issues

We first analyze the impact of incentive issues on the optimal decisions of the firm along with the social welfare and consumer surplus for a fixed pair of $d_0$ and $\nu$. To do so, we introduce a comparison benchmark, the first-best setting, where the firm does not face any incentive issues.

## 4.1 First-Best Benchmark

In the first-best setting, the firm has the capability to develop the algorithm and does not face any incentive issues. The first-best problem then corresponds to Problem (6) without incentive constraints (7) and (9) except that the firm directly incurs cost $\kappa$. Specifically, the problem without incentive issues, denoted by $J^{\mathtt{FB}}_{1a}(d_0)$, is obtained by setting payments to $\kappa$ (i.e., $x_{1h} = x_{1\ell} = x_{2h} = x_{2\ell} = \kappa$) and removing incentive constraints (7) and (9) in Problem (6). The following proposition characterizes the firm's optimal decision at first-best.

PROPOSITION 2. *The optimal solution to* $J^{\mathtt{FB}}_{1a}(d_0)$ *is unique such that* $\delta_{1h} = \delta^{\mathtt{M}}$ *and* $\delta_{1\ell} = \delta^{\mathtt{FB}}$ *where* $\delta^{\mathtt{FB}}$ *is the unique solution* $\max_{\delta\in[0,\nu]} R_\nu(\delta, \alpha_\ell) + \pi_w(d_0+\delta)(\alpha_h - \alpha_\ell)\tau$, *and* $\delta^{\mathtt{FB}} > \delta^{\mathtt{M}}$.

We denote by $\delta^{\mathtt{FB}}$ the optimal value of $\delta_{1\ell}$ maximizing $J^{\mathtt{FB}}_{1a}(d_0)$. We also refer to $p^{\mathtt{FB}}$ as the corresponding price that yields data size $\delta^{\mathtt{FB}}$. If the firm succeeds in developing the initial algorithm with high accuracy ($\alpha_1 = \alpha_h$), no further improvement is necessary, and the firm charges the optimal myopic price $p^{\mathtt{M}}$, inducing $\delta^{\mathtt{M}}$ over the remaining time horizon. If this accuracy is low ($\alpha_1 = \alpha_\ell$),

however, the firm faces a tradeoff between maximizing revenues in the current period or acquiring additional data to leverage the AI Flywheel effect. In this case, the firm underprices with $p^{\mathtt{FB}} < p^{\mathtt{M}}$ and forfeits the optimal myopic revenue to collect more data, that is, $\delta^{\mathtt{FB}} > \delta^{\mathtt{M}}$, increasing the probability of high accuracy and hence expected profit in the next period.

To the extent that we consider a dynamic model, these distortions away from the myopic decision ($\delta^{\mathtt{FB}} > \delta^{\mathtt{M}}$) are to be expected. They occur in many different settings, such as in inventory problems (see Heese and Swaminathan 2010, for example), but are especially present in problems of pricing under positive adoption externalities (see Section 2).

Finally, the firm's pricing and data collection decisions $p^{\mathtt{FB}}$ and $\delta^{\mathtt{FB}}$ determine the social welfare and customer surplus at first best, which we denote by $W^{\mathtt{FB}}$ and $C^{\mathtt{FB}}$, respectively. Customer surplus $C^{\mathtt{FB}}$ is the total valuations net of the product price of all customers who make a purchase. Social welfare $W^{\mathtt{FB}}$ is then the sum of customer surplus $C^{\mathtt{FB}}$ and the firm's profit net of effort cost $\kappa$.

### 4.2 Overpricing, Underpricing, and Optimal Data Collection

To alleviate the moral hazard issue, the firm may need to incur costly deviations away from first-best decisions. The next result shows that the monotonicity of data impact $\rho(\cdot)$ is sufficient to determine when the firm overprices and when it underprices.

THEOREM 1. *We have*

1. *if $\rho(\cdot)$ is constant, then $\delta^* = \delta^{\mathtt{FB}}$,*
2. *if $\rho(\cdot)$ is strictly increasing, then $\delta^* > \delta^{\mathtt{FB}}$,*
3. *if $\rho(\cdot)$ is strictly decreasing, then $\delta^* < \delta^{\mathtt{FB}}$.*

Theorem 1 shows that the monotonicity of the data impact induces the firm to deviate from first best in different directions. In essence, as the availability of data increases, the performances of a naive and a specialized algorithm are improved differently, which then affects returns to effort in the contract. To compensate this issue, the firm may need to incur costly deviations away from the first-best decisions. In this sense, the increase/decrease (with respect to $\delta^{\mathtt{FB}}$) in collected data characterized by Theorem 1 resembles deviations from first best found in other contexts (such as over/understocking decisions in, e.g., Alles et al. 1995, Dai and Jerath 2013, 2019 and de Véricourt and Gromb 2018, 2019).

Overall, Theorem 1 defines three distinct regimes. First, when the relative impact of shirking is independent of the dataset size ($\rho$ is constant), then no deviation from first best is required.

However, if the data impact is increasing, the net effect of shirking increases faster than the probability of high accuracy. In this case, high accuracy is more indicative of efforts at higher data volumes, and the rent is decreasing in data size $\delta$. Thus, the firm underprices with $p^* < p^{\mathtt{FB}}$ in order

to collect more data with $\delta^* > \delta^{\mathtt{FB}}$. In this sense, the incentive issues induce the firm to leverage the AI Flywheel effect even more. In particular, the expected revenue in the first period is lower than first best due to incentive issues, but the expected revenues in the second period are higher since, with more data, the probability of high accuracy increases.

Finally, if the data impact is decreasing, high accuracy is less indicative of efforts at higher data levels, and the rent is increasing in data size $\delta$. Thus, the firm overprices with $p^* > p^{\mathtt{FB}}$ so as to collect less data with $\delta^* < \delta^{\mathtt{FB}}$. The firm's expected revenues in the last period decrease compared to first best since the probability of high accuracy is lower. The revenues in the first period, however, might actually increase compared to first best. This situation happens, for instance, when $p^{\mathtt{M}} > p^* > p^{\mathtt{FB}}$. Overall, the incentive issues prevent the firm from fully leveraging the AI Flywheel effect in this regime.

### 4.3 Social Welfare and Customer Surplus

Theorem 1 uncovers effects that have further implications for social welfare and customer surplus in the presence of incentive issues, which we denote by $W^*$ and $C^*$, respectively. In this case, the social welfare corresponds to the sum of the firm's profit, customer surplus $C^*$, and the provider's profit.[11]

We next characterize how the incentive issues affect the social welfare and customer surplus. In fact, the following result indicates that incentive issues may *increase* these benefits.

Theorem 2. *We have*

1. *if $\rho(\cdot)$ is constant, then $W^* = W^{\mathtt{FB}}$ and $C^* = C^{\mathtt{FB}}$,*
2. *if $\rho(\cdot)$ is strictly increasing, then $W^* > W^{\mathtt{FB}}$, and $C^* > C^{\mathtt{FB}}$*
3. *if $\rho(\cdot)$ is strictly decreasing, then $W^* < W^{\mathtt{FB}}$, and $C^* < C^{\mathtt{FB}}$.*

In other words, the firm's distortion from the first-best benchmark to mitigate the effects of incentive issues improves the social welfare and the customer surplus when the data impact is increasing. By contrast, this distortion hurts the social welfare when the data impact is decreasing.

This result stems from the distortions characterized in Theorem 1 and the fact that the social benefits are monotone in the firm's data collection decision at equilibrium (see Lemma 3 in Appendix B). That a higher $\delta^*$ should improve the social benefits at equilibrium is not obvious a priori. In particular, a larger data collection corresponds to a higher expected accuracy in the next period. This, however, also increases the expected equilibrium price in Period 2b (see Lemma 1 in Appendix A). Our result shows that the latter effect is dominated by the former, and thus both the social welfare and the customer surplus increase with data collection $\delta^*$.

[11] This definition of the social welfare also corresponds to the sum of customer surplus and the firm's profit net of effort cost $\kappa$, which is provided in Section 4.1. Indeed, the transfers between the firm and the provider cancel each other out.

# 5 Impact of the Data Acquisition Capacity

One key product attribute that firms commonly seek to improve is the amount of data that the product collects on its user. In our setup, this corresponds to increasing data acquisition capacity $\nu$. In the following, we explore how this improvement affects the firm's data collection strategy and expected profit. The main finding of this analysis is that data acquisition capacity $\nu$ can have a non-monotonic impact on the firm's profit.

## 5.1 Optimal Data Collection

Without incentive issues, a higher data acquisition capacity induces the firm to collect even more data under the AI Flywheel effect, as seen in the next result.

PROPOSITION 3. *The first-best data size $\delta^{\mathsf{FB}}$ is increasing in data acquisition capacity $\nu$.*

If the firm considers designing a product that acquires more data per user, the firm collects in aggregate even more data ($\delta^{\mathsf{FB}}$ is nondecreasing in $\nu$). Recall that $\delta = \nu q$ and thus increasing $\nu$ provides an opportunity to increase data size $\delta$ to improve the algorithm in the next period while maintaining quantity $q$ and hence revenues in the current period.

The need to mitigate incentive issues, however, sometimes reverses this effect and pushes the firm to collect *less* data when the product's data acquisition capacity is higher. The next theorem, one of our main results, formalizes this finding.

THEOREM 3. *We have the following:*

1. *If $\rho(\cdot)$ is strictly increasing, $\delta^*$ is strictly increasing in $\nu$.*
2. *If $\rho(\cdot)$ is strictly decreasing, a unique threshold $\hat{\nu}$ exists such that $\delta^*$ is strictly increasing in $\nu$ if $\nu \leq \hat{\nu}$ and is strictly decreasing otherwise.*

Thus, increasing the data acquisition capacity induces the firm to collect more data overall when the data impact is also increasing, i.e., when more data mitigates the incentive issues. Intriguingly, this effect may continue to hold when the data impact is decreasing, i.e., when more data actually exacerbates the incentive issues. This happens as long as the data acquisition capacity is not too large ($\nu \leq \hat{\nu}$). When the data acquisition capacity becomes large ($\nu > \hat{\nu}$), acquiring more data per user actually induces the firm to collect less data overall (see Proposition EC.6 in Section EC.3 of the e-companion for general conditions ensuring $\hat{\nu} < \bar{\nu}$).

More specifically, when the acquisition capacity increases, the firm has an incentive to collect more data to benefit more from the AI Flywheel effect, as discussed in the first-best benchmark (see Proposition 3). When the data impact is increasing, collecting more data also reduces the intensity of the moral hazard problem and hence the agency costs in the next period. Both effects are aligned in this case, and the firm increases $\delta^*$ as a result. By contrast, collecting more data

intensifies the moral hazard problem in the next period when the data impact is decreasing. The firm then faces a tradeoff between leveraging the AI Flywheel effect or reducing future agency costs. When the acquisition capacity is small, the amount of collected data $\delta = \nu q$ remains small for any quantity $q$, which yields low agency costs. As the acquisition capacity increases, boosting the AI Flywheel effect dominates the increases in agency costs, and data size $\delta^*$ increases as a result. When the acquisition capacity becomes large enough (i.e., when $\nu > \hat{\nu}$), however, the agency costs dominates the revenues due to the AI Flywheel effect. The firm then focuses on reducing these agency costs by decreasing $\delta^*$.

Note finally that the data collection strategies characterized in Theorem 3 do not fully determine the firm's pricing decisions. When the firm collects less data even though it improves its data acquisition capacity, the firm charges a higher price $p^*$ so as to decrease $q^*$ with $\delta^* = \nu q^*$. However, when the firm collects more data as the acquisition capacity increases, the firm may either increase or decrease $p^*$ depending on the problem parameters (and the third-order behavior of the learning curves in particular). This is because an increase in $\delta^* = \nu q^*$ does not require $q^*$ to increase when $\nu$ increases.

### 5.2 The Firm's Profit

To the extent that acquisition capacity $\nu$ determines the firm's data collection strategy per Theorem 3, improving this capability ultimately affects the firm's profit. The next result – one of the main findings of the paper – characterizes this effect.

THEOREM 4. *We have:*

1. *If $\rho(\cdot)$ is strictly increasing, the expected profit of the firm is strictly increasing in $\nu$.*
2. *If $\rho(\cdot)$ is strictly decreasing, a unique threshold $\tilde{\nu}$ exists such that the expected profit is strictly increasing in $\nu$ if $\nu \leq \tilde{\nu}$ and is strictly decreasing otherwise. Further, threshold $\tilde{d} > 0$ exists such that the profit is strictly decreasing (i.e., $\tilde{\nu} = 0$) if $d_0 \geq \tilde{d}$.*

Thus, when the data impact is increasing, the firm's profit is higher when the product can acquire more data on its user. In fact, this effect continues to hold even when more data exacerbates the incentive issues, as long as the data acquisition capacity is not too large ($\nu \leq \tilde{\nu}$).

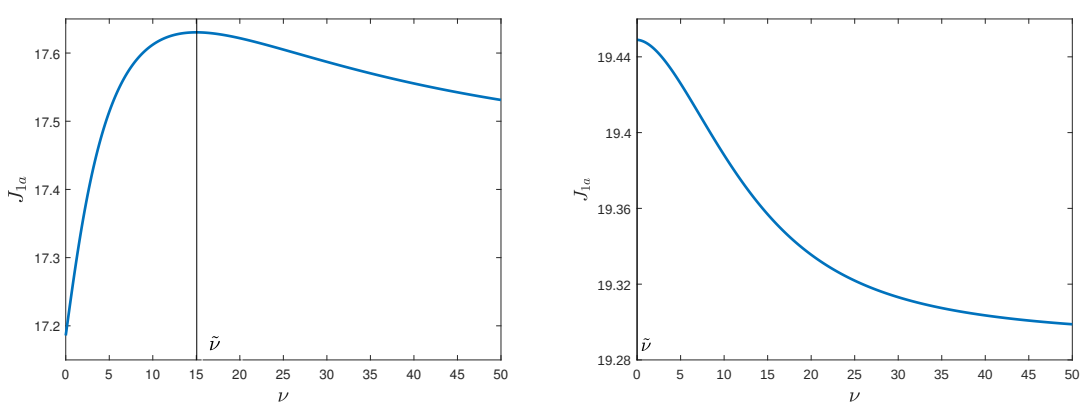
However, if the data acquisition capacity is large enough ($\nu > \tilde{\nu}$), boosting the product's capacity to collect data may actually hurt the firm's profit. Indeed, the second part of Theorem 4 states that the profit has a unique maximum in $\nu$. For certain problem parameters (and learning curves in particular), this maximum is sometimes achieved at $\tilde{\nu} = \bar{\nu}$. We provide general conditions for $\tilde{\nu} < \bar{\nu}$ in the e-companion (see Proposition EC.6 in Section EC.3). These conditions hold when $\pi_w(d) - \pi_s(d)$ is decreasing in $d$ and converges to 0 as $d$ approaches infinity (see Proposition EC.3

in Section EC.2 of the e-companion), a property that the learning curves used in the ML literature typically satisfy (see Section EC.2 of the e-companion for specific examples).

Three main components of our model jointly shape the impact of the data acquisition capacity on the firm's profit: the AI Flywheel effect, pricing for revenue, and incentive issues. To better leverage the AI Flywheel effect, the firm needs to collect more data. Therefore, optimal data collection $\delta^*$ deviates from myopic data collection $\delta^{\mathtt{M}}$, and the optimal price $p^*$ deviates from the myopic price $p^{\mathtt{M}}$ (see Section 4.2), which hurts revenues in the first period. Improving the data acquisition capacity $\nu$ makes it possible for the firm to push the price toward the myopic one, thus boosting revenues in the current period without affecting the amount of data collected to retrain the algorithm in the next period. This positive effect continues to hold for larger values of the acquisition capacity when the data impact is increasing (per the first item of Theorem 4).

When the data impact is decreasing, however, the previous positive effect holds until the acquisition capacity hits threshold $\tilde{\nu}$, which is exactly the value at which the firm charges the myopic price ($p^* = p^{\mathtt{M}}$). In this case, the firm has maximized the revenue it can collect from pricing in the first period. Increasing the data acquisition capacity further exacerbates the incentive issues in the second one. In addition, the marginal contribution to the accuracy of the additional collected data is decreasing (recall that $\pi_w$ is increasing but concave). Thus, exceeding $\tilde{\nu}$ generates no additional revenue in the current period, while the improvements in accuracy and hence revenues in the second one are offset by the increase in agency costs. The firm's profit decreases as a result. Figure 2a

**Figure 2** **Effect of the product's data acquisition capacity on the firm's profit when more data exacerbate the incentive issues**




(a) $d_0 < \tilde{d}$ (b) $d_0 \geq \tilde{d}$

*Note.* Probabilities generating this example are $\pi_w(d) = 1 - 1/\sqrt{d}$, $\pi_s(d) = \pi_w(d)(1 - 1/(d+1))$ and the data impact is $\rho(d) = 1/(1+d)$, which is decreasing. The value of $\tilde{d} \approxeq 10$.

illustrates this last point and depicts the effect of $\nu$ on the firm's expected profit $J_{1a}$ when the data

impact is decreasing. The figure shows that the data acquisition capacity $\nu$ has a unimodal effect on profit. Threshold $\hat{\nu}$ corresponds to the straight black line in Figure 2a, with $\tilde{\nu} \approxeq 15$.

This unimodal relationship between the data acquisition capacity and the firm's profit has further implications for the role of initial data size $d_0$. Indeed, when the initial dataset of the firm is large enough, increasing the product's acquisition capacity always hurts profits when the data impact is decreasing. Because the initial dataset is large, the firm needs to rely less on the AI Flywheel effect to collect data for the second period. The firm nonetheless charges a price above the myopic one to mitigate the incentive issues. If the data acquisition capacity improves, the incentive issues intensify, and the firm increases the price away from the myopic one, which decreases revenues. Figure 2b depicts an example of a sufficiently large initial dataset ($d_0 \geq \tilde{d} \approxeq 10$) and shows that the firm's profit strictly decreases in the data acquisition capacity.

# 6 Extensions

In this section, we explore the robustness of our results – and Theorem 4 in particular – in a variety of settings that extend our basic setup. We consider next different setups where the firm withholds data, offers short-term contracts or lets the provider shirk.

## 6.1 Withholding Data

In some instances, the firm may be able to withhold a portion of the collected data from the provider.[12] In this case, the firm needs to decide how much data to share, which in turn affects the incentive issues.

When data impact $\rho$ is increasing, more data mitigates the incentive issues and boosts the AI Flywheel effect. Hence, the firm has no incentive to withhold data from the provider. By contrast, more data exacerbates the incentive issues when the data impact is decreasing. As a result, the firm may still collect data $\delta$ to boost its current sales revenues but only share $\lambda\delta$, for some $\lambda \in [0, 1]$, to reduce the agency costs. This, however, also hurts the AI Flywheel effect, and the firm needs to jointly optimize over $\delta$ and $\lambda$ to make these tradeoffs.

Thus, withholding data may change our findings and in particular the effect of $\nu$ on the firm's profit. The next result, however, shows that a weaker version of Theorem 4 holds in this setting.

Theorem 5. *Assume the firm can withhold a portion of its collected data. We have:*

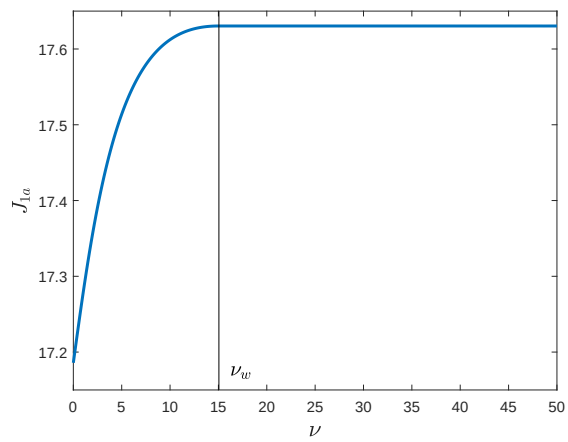
1. *If $\rho$ is increasing, the expected profit of the firm is increasing in $\nu$.*

[12] Note that most ML service providers are also Cloud Service Providers (CSP). Indeed, user data from digital products are typically collected and stored in the cloud by CSPs, which take advantage of this to offer AI services (see, e.g., MSV 2018). This makes it hard for their clients to withhold data.

2. *If $\rho$ is decreasing, a unique threshold $\nu_w$ exists such that the expected profit is increasing in $\nu$ if $\nu \leq \nu_w$ and is constant otherwise. Further, threshold $d_w > 0$ exists such that the profit is constant (i.e., $\nu_w = 0$) if $d_0 \geq d_w$.*

This results first shows that Theorem 4 continues to hold with data withholding when $\rho$ increases. By contrast, if $\rho$ decreases, boosting the data acquisition capacity above a threshold does not improve the firm's profit. This is because withholding data decreases the agency costs, which compensates for the decrease in profit characterized in Theorem 4. Figure 3 illustrates this finding, with $\nu_w \approxeq 15$. Because improving the acquisition capacity is not free in general, the overall insight of Section 5.2 continues to hold in this setting.

**Figure 3 Effect of the product's data acquisition capacity on the firm's profit when more data exacerbate the incentive issues and the firm can withhold data from the provider**




*Note.* Probabilities generating this example are $\pi_w(d) = 1 - 1/\sqrt{d}$, $\pi_s(d) = \pi_w(d)(1 - 1/(d+1))$, and the data impact is $\rho(d) = 1/(1+d)$, which is decreasing. The size of the initial dataset is $d_0 = 5 < d_w \approxeq 10$.

## 6.2 Competitive Provider Market and Short-term Contracting

In our basic setup, the firm offers a long-term contract with a unique provider. When the market of providers is competitive, however, the firm is better off terminating the contract when accuracy is low in the first period and hiring a new provider for the second one. Indeed, the threat of firing the provider reduces the information rent. Nonetheless, the firm's pricing and data collection in this first period continue to influence the incentive issues with the next provider. We find that all our results hold in this setting as well (see Section EC.4 of the e-companion for the formal results). The only difference is the distortions away from first best, which are milder in this case due to the termination threat (see Lemma EC.2 in the e-companion). Note also that re-contracting with the same provider in the second period is suboptimal for the firm because the termination threat

disappears. If the firm contracts with a single provider over both periods, it prefers a long-term contract to short-term ones. Indeed, with short-term contracts, the firm adjusts its data collection strategy to maximize the continuation profit without accounting for the agency costs of the first period. In this case, the effect of the acquisition capacity in Theorem 4 holds for the continuation profit (see Section EC.5 of the e-companion).

### 6.3 Allowing Shirking

We have considered thus far situations where the provider works in both periods. In other contexts, it may be preferable for the firm to let the provider shirk. This may in turn change the effect characterized in Theorem 4. Indeed, we show in the e-companion (see Proposition EC.14 in Section EC.7) that when the provider shirks in the second period, the profit always increases in the acquisition capacity. This is because the collected data in the first period do not interact with incentive issues in the second one. In this sense, the problem becomes analogous to our first-best benchmark, albeit with a milder AI Flywheel effect since $\pi_s(d) < \pi_w(d)$. However, when the provider shirks in the first period but works in the second one, the amount of data collected in the first period affects the agency costs across periods. In this case, we show that all of our findings hold (see Proposition EC.15 in Section EC.7 of the e-companion).

## 7 Conclusion

This paper proposes a simple dynamic framework to study how firms that outsource the development of their ML algorithm can leverage the AI Flywheel effect. Our setup accounts for the three main features of this problem: i) the tradeoff between improving the algorithms' accuracy and maximizing revenues due to the AI Flywheel effect, ii) the need to manage the incentive issues that outsourcing the algorithm brings about, and iii) the interaction between the amount of data on which the algorithm is trained and the efficacy of the provider's effort. We further introduce the notion of data impact as a framework to represent the interaction between data and effort.

Taken together, our results identify three different regimes which depend on the nature of the data impact. These regimes determine whether the firm overprices or underprices, and regulate the impact of the data acquisition capacity on the firm's profit and decisions. In particular, when the data impact decreases, we find that improving the capacity to acquire user data may hurt profits and induce the firm to actually collect less data overall. These regimes have further implications for society. Notably, incentive issues improve social welfare when the data impact is increasing.

Improving a product's capacity to acquire usage data is a key design choice in practice but can be costly and at times challenging (e.g., Boonstra et al. 2018). As such, our insights on when a firm should seek to improve this acquisition capacity address an important managerial problem.

In essence, the acquisition capacity determines the strength of the positive adoption externalities of the AI Flywheel effect in our setup. In the context of classical products, this corresponds to the intensity of the customer social network or how fast information circulates in the market (Shy 2011), which the firm cannot easily influence. Nonetheless, the stronger the network effect in classical adoption settings, the better off the firm typically is (Gabszewicz and Garcia 2008). To some extent, this remains true for data products. In our setup, entrepreneurs who seek to leverage the AI Flywheel effect should indeed consider enhancing the data collection capability of their products, but only up to a certain level. Allowing a data product to collect too much usage data may sometimes damage the firm's profits.

Our results further provide predictions that future work can empirically test. In particular, given that the existing literature points to the importance of collecting a large amount of data over improving algorithms (Banko and Brill 2001, Halevy et al. 2009), we expect the data impact to decrease in many practical contexts. Our work thus provides theoretical support for the hypothesis that firms set higher prices for a new AI product when the product's algorithm is outsourced (per Theorem 1). In addition, we predict that a significant increase in the product's capacity to generate usage data induces firms to collect less data to improve the algorithm, which hurts their profits (per Theorem 3 and Theorem 4, respectively).

From a technical perspective, our model can be extended in a variety of ways. For instance, exploring how information asymmetry regarding the provider's skill interacts with the AI Flywheel effect constitutes a fruitful future research direction. Similarly, future research could explore our setup with a longer time horizon. Another interesting direction of future research is to explore in more detail when and in which of the two periods shirking is optimal. In this case, an intriguing question is whether the firm should incentivize high effort or instead focus on increasing demand in the first period.

More generally, we believe that our work opens up more fundamental research directions and questions for the management of data-driven business models. Specifically, our paper considers a problem in which the provision of data interacts with incentive issues. Indeed, the key aspect of our setup is that the principal can regulate the intensity of the moral hazard problem she faces by controlling (through pricing in our setting) the data to which the agent has access. We believe that this interaction between data and incentives is present in many other contexts than the outsourcing of the AI Flywheel effect.

Overall, our results provide the first insights on how firms can leverage the AI Flywheel effect. In addition, ours is the first paper to consider the problem of contracting ML algorithms. Given the shortage of data scientists and the growing outsourcing industry in this domain, we expect the issue to gain importance in the coming years.

## References

Acemoglu D, Makhdoumi A, Malekian A, Ozdaglar A (2019) Too much data: Prices and inefficiencies in data markets. Technical report, National Bureau of Economic Research.

Aflaki A, Feldman P, Swinney R (2019) Becoming strategic: Endogenous consumer time preferences and multiperiod pricing. *Forthcoming, Operations Research* .

Alles M, Datar SM, Lambert RA (1995) Moral hazard and management control in just-in-time settings. *Journal of Accounting Research* 33:177–204.

Anderson Jr EG, Parker GG (2013) Integration and cospecialization of emerging complementary technologies by startups. *Production and Operations Management* 22(6):1356–1373.

Banko M, Brill E (2001) Scaling to very very large corpora for natural language disambiguation. *Proceedings of the 39th annual meeting on association for computational linguistics*, 26–33 (Association for Computational Linguistics).

Bensaid B, Lesne JP (1996) Dynamic monopoly pricing with network externalities. *International Journal of Industrial Organization* 14(6):837–855.

Bimpikis K, Crapis D, Tahbaz-Salehi A (2019) Information sale and competition. *Management Science* 65(6):2646–2664.

Blagec K, Dorffner G, Moradi M, Samwald M (2020) A critical analysis of metrics used for measuring progress in artificial intelligence. *arXiv preprint arXiv:2008.02577* .

Boonstra TW, Nicholas J, Wong QJ, Shaw F, Townsend S, Christensen H (2018) Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges and potential solutions. *Journal of medical Internet research* 20(7):e10131.

Bursztein E (2018) Attacks against machine learning ? an overview. `elie.net/blog/ai/attacks-against-machine-learning-an-overview/` [Online; accessed 27-January-2020].

Cabral LM, Salant DJ, Woroch GA (1999) Monopoly pricing with network externalities. *International Journal of Industrial Organization* 17(2):199–214.

Campbell A (2013) Word-of-mouth communication and percolation in social networks. *American Economic Review* 103(6):2466–98.

Cengiz K (2015) 10 startups that have become a success with outsourced development. `https://perfectial.com/blog/?s=10+Startups+That+Have+Become+A+Success+With+Outsourced+Development&search-type=normal` [Online; accessed 17-January-2020].

Choudhary V, Ghose A, Mukhopadhyay T, Rajan U (2005) Personalized pricing and quality differentiation. *Management Science* 51(7):1120–1130.

Dai T, Jerath K (2013) Salesforce compensation with inventory considerations. *Management Science* 59(11):2490–2501.

Dai T, Jerath K (2019) Salesforce contracting under uncertain demand and supply: Double moral hazard and optimality of smooth contracts. *Marketing Science, Forthcoming* .

de Véricourt F, Gromb D (2018) Financing capacity investment under demand uncertainty: An optimal contracting approach. *Manufacturing & Service Operations Management* 20(1):85–96.

de Véricourt F, Gromb D (2019) Financing capacity with stealing and shirking. *Management Science* 65(11):5128–5141.

Deshpande I (2019) What is customer data? definition, types, collection, validation and analysis. `https://www.martechadvisor.com/articles/data-management/customer-data-definition-types-collection-validation-analysis-martech101/` [Online; accessed 10-March-2020].

Feldman P, Papanastasiou Y, Segev E (2018) Social learning and the design of new experience goods. *Management Science* 65(4):1502–1519.

Ferri C, Hernández-Orallo J, Modroiu R (2009) An experimental comparison of performance measures for classification. *Pattern Recognition Letters* 30(1):27–38.

Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH (2012) Predicting sample size required for classification performance. *BMC medical informatics and decision making* 12(1):1–10.

Gabszewicz JJ, Garcia F (2008) A note on expanding networks and monopoly pricing. *Economics Letters* 98(1):9–15.

Golden K (2017) Deere to advance machine learning capabilities in acquisition of blue river technology. `www.deere.com/en/our-company/news-and-announcements/news-releases/2017/corporate/2017sep06-blue-river-technology/` [Online; accessed 27-January-2020].

Gromb D, Martimort D (2007) Collusion and the organization of delegated expertise. *Journal of economic theory* 137(1):271–299.

Gu B, Hu F, Liu H (2001) Modelling classification performance for large data sets. *International Conference on Web-Age Information Management*, 317–328 (Springer).

Gutierrez D (2016) Data-first machine learning. `http://www.inetservicescloud.com/data-first-machine-learning/` [Online; accessed 11-September-2020].

Hackathorn R (2018) How managers should prepare for deep learning: New paradigms. `https://towardsdatascience.com/how-managers-should-prepare-for-deep-learning-new-paradigms-28de63054ea6` [Online; accessed 15-09-2020].

Halevy A, Norvig P, Pereira F (2009) The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems* 24(2):8–12, ISSN 1541-1672, URL `http://dx.doi.org/10.1109/MIS.2009.36`.

Heese HS, Swaminathan JM (2010) Inventory and sales effort management under unobservable lost sales. *European Journal of Operational Research* 207(3):1263–1268.

Jo J, Bengio Y (2017) Measuring the tendency of cnns to learn surface statistical regularities. *arXiv preprint arXiv:1711.11561* .

Katz ML, Shapiro C (1994) Systems competition and network effects. *Journal of economic perspectives* 8(2):93–115.

Kendra A, Kumar S, O'Brien D, Shankar R, Snover J, Viljoen S (2019) Failure modes in machine learning. `https://docs.microsoft.com/en-us/security/failure-modes-in-machine-learning` [Online; accessed 17-January-2020].

Kos T, Kosar T, Mernik M (2012) Development of data acquisition systems by using a domain-specific modeling language. *Computers in industry* 63(3):181–192.

Kräkel M, Schöttner A (2016) Optimal sales force compensation. *Journal of Economic Behavior & Organization* 126:179–195.

Laffont JJ, Martimort D (2009) *The theory of incentives: the principal-agent model* (Princeton university press).

Leite R, Brazdil P (2005) Predicting relative performance of classifiers from samples. *Proceedings of the 22nd international conference on machine learning*, 497–503.

Lipton ZC (2016) The mythos of model interpretability. *arXiv preprint arXiv:1606.03490* .

Loertscher S, Marx LM (2019) Digital monopolies: Privacy protection or price regulation? .

McGrath MJ, Scanaill CN (2013) Key sensor technology components: hardware and software overview. *Sensor Technologies*, 51–77 (Springer).

Mehta S, Dawande M, Janakiraman G, Mookerjee V (2019) How to sell a dataset? pricing policies for data monetization. *Pricing Policies for Data Monetization (February 12, 2019)* .

Miller R (2016) At amazon the flywheel effect drives innovation. `techcrunch.com/2016/09/10/at-amazon-the-flywheel-effect-drives-innovation/` [Online; accessed 27-January-2020].

MSV J (2018) The rise of artificial intelligence as a service in the public cloud. `https://www.forbes.com/sites/janakirammsv/2018/02/22/the-rise-of-artificial-intelligence-as-a-service-in-the-public-cloud/?sh=12e5ed0b198e` [Online; accessed 07-May-2021].

Ng A (2018) Ai transformation playbook. `www.coursera.org/learn/ai-for-everyone/lecture/kSWz6/ai-transformation-playbook-part-2` [Online; accessed 27-January-2020].

Nicolaus Henke J Bughin, Chui M, Manyika J, Saleh T, Wiseman B, Sethupathy G (2016) The age of analytics: competing in a data-driven world.

Perlich C, Provost F, Simonoff J (2003) Tree induction vs. logistic regression: A learning-curve analysis. *Journal of Machine Learning Research* 211–255.

Research Nester Pvt Ltd (2019) Data analytics outsourcing market.

Ribeiro MT, Singh S, Guestrin C (2016) Why should i trust you?: Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135–1144 (ACM).

Sarikaya R (2019) How alexa learns. `https://blogs.scientificamerican.com/observations/how-alexa-learns/` [Online; accessed 27-January-2020].

Schmitz PW (2005) Allocating control in agency problems with limited liability and sequential hidden actions. *RAND Journal of Economics* 318–336.

Schmitz PW (2013) Job design with conflicting tasks reconsidered. *European Economic Review* 57:108–117.

Schöttner A (2016) Optimal sales force compensation in dynamic settings: Commissions vs. bonuses. *Management Science* 63(5):1529–1544.

Shy O (2011) A short survey of network economics. *Review of Industrial Organization* 38(2):119–149.

Sokolova M, Japkowicz N, Szpakowicz S (2006) Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. *Australasian joint conference on artificial intelligence*, 1015–1021 (Springer).

Spencer CD (1990) *Digital design for computer data acquisition* (Cambridge University Press).

Swinney R, Cachon GP, Netessine S (2011) Capacity investment timing by start-ups and established firms in new markets. *Management Science* 57(4):763–777.

Tanrısever F, Erzurumlu SS, Joglekar N (2012) Production, process investment, and the survival of debt-financed startup firms. *Production and Operations Management* 21(4):637–652.

Tharwat A (2020) Classification assessment methods. *Applied Computing and Informatics* .

Trautman E (2018) The virtuous cycle of ai products. `https://www.eriktrautman.com/posts/the-virtuous-cycle-of-ai-products` [Online; accessed 27-January-2020].

Viering T, Loog M (2021) The shape of learning curves: a review. *arXiv preprint arXiv:2103.10948* .

Vincent J (2016) Twitter taught microsoft's ai chatbot to be a racist asshole in less than a day. `https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist` [Online; accessed 11-September-2020].

Yu M, Debo L, Kapuscinski R (2015) Strategic waiting for consumer-generated quality information: Dynamic pricing of new experience goods. *Management Science* 62(2):410–435.

# Appendix

## A. Optimal Decisions of the Firm

We characterize the optimal decisions of the firm when it needs to outsource the development of the algorithm. These decisions correspond to the amount of collected data $\delta_{1\ell}, \delta_{1h}$ (and the corresponding price $p_{1\ell}, p_{1h}$) as well as all payments $x_{1\ell}, x_{1h}, x_{2\ell}$, and $x_{2h}$ in $J_{1a}$.

The optimization problem of $J_{1a}$ in (6) corresponds to maximizing the expected profit, that is, the total revenue net of the payments motivating the provider to exert effort. This problem can be solved in three steps. Starting from Period 2b, we initially consider the myopic problem $J_{2b}$, whose solution is characterized in Lemma 1. Next, fixing the amount of collected data $\delta_{1\ell}$ and $\delta_{1h}$, we characterize the corresponding payments. The following lemma (Lemma 2) provides the optimal payments that motivate the exertion of effort for given $\delta_{1\ell}, \delta_{1h}$. Finally, we characterize the optimal amount of collected data that maximizes the total expected profit of the firm.

Specifically, the firm's problem in Period 2b corresponds to the myopic problem in (2). Straightforward calculations then lead to the following result.

Lemma 1. *Given accuracy $\alpha_2$ and dataset size $d_1$, the optimal collected data size $\delta^{\mathtt{M}}$ and expected profit $J_{2b}(\alpha_2)$ are equal to $\delta^{\mathtt{M}} = \nu \bar{F}(\phi^{-1}(0))$ and $J_{2b}(\alpha_2) = \alpha_2 \tau$, respectively, where $\tau \triangleq \phi^{-1}(0)\bar{F}(\phi^{-1}(0))$ and $\phi^{-1}(\cdot)$ is the inverse of virtual value function $\phi(\cdot)$.*

The proof of Lemma 1 is provided alongside selected proofs in Appendix B. The corresponding myopic price that yields data size $\delta^{\mathtt{M}}$ is then equal to, per Section 3, $p^{\mathtt{M}} = \alpha_2 \phi^{-1}(0)$, and quantity $\tau$ is equal to $p^{\mathtt{M}}(\delta^{\mathtt{M}}/\nu)/\alpha_2$, which is the marginal revenue per unit of accuracy under optimal myopic pricing. Note that $\delta^{\mathtt{M}}$ does not depend on accuracy, but the optimal price $p^{\mathtt{M}}$ and profit $J_{2b}$ do.

Lemma 2. *Fix $\delta_{1\ell}$ and $\delta_{1h}$ in (6); then the corresponding optimal payments are $x_{1\ell}(\delta_{1\ell}) = x_{2\ell}(\delta_{1\ell}) = 0$, and*

$$x_{1h}(\delta_{1\ell}) = \frac{\kappa}{\pi_w(d_0) - \pi_s(d_0)} + \left(\frac{\kappa}{\rho(d_0 + \delta_{1\ell})} - \kappa\right) \text{ and } x_{2h}(\delta_{1\ell}) = \frac{\kappa}{\pi_w(d_0 + \delta_{1\ell}) - \pi_s(d_0 + \delta_{1\ell})}.$$

*The firm's total expected payment to the provider is then given by $\kappa/\rho(d_0) + \kappa/\rho(d_0 + \delta_{1\ell})$.*

In the first period, the firm needs to account for the provider's future expected profits contrary to the payments in Period 2. Specifically, the optimal payment $x_{1h}(\delta_{1\ell})$ corresponds to i) the bonus payment taken at $d_0$ augmented by ii) the provider's rent of the second period net of effort cost because, in our setup, the firm cannot easily replace the provider across periods. In this sense, the second term of $x_{1h}(\delta_{1\ell})$ captures the cost due to the scarcity of AI service providers in the market. In both periods, the ex-post payments to the provider depend on the amount of data $\delta_{1\ell}$. Because

there is no need for further improvement after a high accuracy $\alpha_h$ in the first period, payments depend only on $\delta_{1\ell}$.

MLRP ensures that under the optimal contract, realizations of higher value ($\alpha_h$ in our setup) are more rewarded ($x_{2h} > x_{2l}$). Thus, MLRP concerns the ex-post payments to the provider. By contrast, Lemma 2 shows that data impact $\rho(\cdot)$ determines the ex-ante provider's rent $\kappa[1/\rho(d_0) + 1/\rho(d_0+\delta_{1\ell})]$.

## B. Proof of Results

*Proof of Lemma 1.* Recall the definition $R_\nu(\delta,\alpha) = \alpha\delta/\nu F^{-1}((1-\delta/\nu))$ in Section 3. $F$ is the c.d.f. of the standard uniform distribution. Therefore, it can be verified that $R_\nu(\delta,\alpha)$ is concave in $\delta$, and the following first-order condition is sufficient for optimality.

$$\frac{1}{\nu}F^{-1}\left(1-\frac{\delta}{\nu}\right) - \frac{\delta}{\nu^2}\frac{1}{f\left(F^{-1}\left(1-\frac{\delta}{\nu}\right)\right)} = 0\,. \tag{10}$$

Let $\xi = F^{-1}\left(1-\frac{\delta}{\nu}\right)$; hence $\delta/\nu = \bar{F}(\xi)$. We first multiply both sides of the equality by $\nu$ and then use this new notation $\xi$. Using the definition of the virtual value function $\phi$, we have $\phi(\xi) = 0$.

Because the virtual value function of the uniform distribution is increasing and crosses 0 at 1/2, we conclude that the unique optimal solution $\delta^{\mathtt{M}}$ to Problem (2) is equal to $\nu\bar{F}(\phi^{-1}(0))$. Evaluating the objective function at the optimal solution and using the fact that $\tau = \phi^{-1}(0)\bar{F}(\phi^{-1}(0))$, we conclude that $J_{2b}(\alpha_2) = \alpha_2\tau$.

*Proof of Lemma 2.* Using Lemma 1 and fixing $\delta_{1\ell}$ and $\delta_{1h}$, we obtain the following problem where we only optimize over the payments $x_{i\ell}$ and $x_{ih}$ for $i = 1,2$.

$$\min_{\substack{x_{1h},x_{1\ell}\geq 0\\ x_{2h},x_{2\ell}\geq 0}} \quad \pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)\left[\pi_w(d_0+\delta_{1\ell})x_{2h} + \bar{\pi}_w(d_0+\delta_{1\ell})x_{2\ell} + x_{1\ell}\right] \tag{11}$$

$$\text{s.t.} \quad \pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)\left[x_{1\ell} + J_p(d_0)\right] - \kappa \geq \pi_s(d_0)x_{1h} + \bar{\pi}_s(d_0)\left[x_{1\ell} + J_p(d_0)\right] \tag{12}$$

$$\pi_w(d_0+\delta_{1\ell})x_{2h} + \bar{\pi}_w(d_0+\delta_{1\ell})x_{2\ell} - \kappa \geq \pi_s(d_0+\delta_{1\ell})x_{2h} + \bar{\pi}_s(d_0+\delta_{1\ell})x_{2\ell} \tag{13}$$

$$J_p(d_0) = \pi_w\left(d_0+\delta_{1\ell}\right)x_{2h} + \bar{\pi}_w\left(d_0+\delta_{1\ell}\right)x_{2\ell} - \kappa. \tag{14}$$

To solve this problem, we first consider a subproblem where we optimize over only $x_{1h}$ and $x_{1\ell}$ for a fixed pair of $x_{2h}$ and $x_{2\ell}$ that satisfy (13). Since $\delta_{1\ell}$ is also fixed, we obtain a constant for $J_p(d_0)$. Thus, the subproblem corresponds to minimizing the expected payments in the first period while satisfying the incentive constraint (12), which is given by

$$\min_{x_{1h},x_{1\ell}\geq 0} \quad \pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)x_{1\ell} \tag{15}$$

$$\text{s.t.} \quad \pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)\left[x_{1\ell} + J_p(d_0)\right] - \kappa \geq \pi_s(d_0)x_{1h} + \bar{\pi}_s(d_0)\left[x_{1\ell} + J_p(d_0)\right] \tag{16}$$

Following Proposition 4.2 in Laffont and Martimort (2009, p. 157), we obtain the optimal solution to the subproblem as $\tilde{x}^*_{1\ell} = 0$ and $\tilde{x}^*_{1h} = \kappa/[\pi_w(d_0) - \pi_s(d_0)] + J_p(d_0)$. Note that this optimal solution

$\tilde{x}^*_{1h}$ and $\tilde{x}^*_{1\ell}$ are for fixed values of $x_{2\ell}$ and $x_{2h}$. Thus, we now optimize over $x_{2\ell}$ and $x_{2h}$ by incorporating the values of $\tilde{x}^*_{1\ell}$ and $\tilde{x}^*_{1h}$ into Problem (11). Arranging terms, we obtain

$$\begin{aligned} \min_{x_{2h},x_{2\ell}\geq 0} \quad & \pi_w\left(d_0+\delta_{1\ell}\right)x_{2h}+\bar{\pi}_w\left(d_0+\delta_{1\ell}\right)x_{2\ell}+\frac{\kappa}{\rho(d_0)}-\pi_w(d_0)\kappa \\ \text{s.t.} \quad & \pi_w(d_0+\delta_{1\ell})x_{2h}+\bar{\pi}_w(d_0+\delta_{1\ell})x_{2\ell}-\kappa\geq \pi_s(d_0+\delta_{1\ell})x_{2h}+\bar{\pi}_s(d_0+\delta_{1\ell})x_{2\ell}\,. \end{aligned} \tag{17}$$

Invoking the same proposition from Laffont and Martimort (2009) one more time, we obtain the optimal solution as $x_{2\ell}(\delta_{1\ell})=0$ and $x_{2h}(\delta_{1\ell})=\kappa/[\pi_w(d_0+\delta_{1\ell})-\pi_s(d_0+\delta_{1\ell})]$. Evaluating $\tilde{x}^*_{1h}$ using $x_{2h}(\delta_{1\ell})$ and $x_{2\ell}(\delta_{1\ell})$, we find $x_{1h}(\delta_{1\ell})=\kappa/[\pi_w(d_0)-\pi_s(d_0)]+\kappa/\rho(d_0+\delta_{1\ell})-\kappa$ and $x_{1\ell}(\delta_{1\ell})=0$. The total expected payment follows from using the values of the optimal solution $x_{ih}$ and $x_{i\ell}$ for $i=1,2$, and we conclude the proof. Q.E.D.

*Proof of Proposition 1.* Using Lemma 1 and the optimal payments derived in Lemma 2, we reformulate Problem (6) as a problem where we only maximize over $\delta_{1\ell}$ and $\delta_{1h}$.

$$\begin{aligned} J_{1a}(d_0)=\max_{\delta_{1h},\delta_{1\ell}\in[0,\nu]} & \pi_w(d_0)\left[R_\nu(\delta_{1h},\alpha_h)+R_\nu(\delta^{\mathtt{M}},\alpha_h)-\frac{\kappa}{\rho(d_0+\delta_{1\ell})}+\kappa\right]-\frac{\kappa}{\rho(d_0)} \\ & +\bar{\pi}_w(d_0)\left[R_\nu(\delta_{1\ell},\alpha_\ell)+\pi_w(d_0+\delta_{1\ell})R_\nu(\delta^{\mathtt{M}},\alpha_h)+\bar{\pi}_w(d_0+\delta_{1\ell})R_\nu(\delta^{\mathtt{M}},\alpha_\ell)-\frac{\kappa}{\rho(d_0+\delta_{1\ell})}\right]. \end{aligned} \tag{18}$$

Note that we can separate the optimization problem above into maximizing over $\delta_{1h}$ and $\delta_{1\ell}$. If we maximize over $\delta_{1h}$, we obtain the myopic problem for which the unique optimal solution is given by Lemma 1 as $\delta_{1h}=\delta^{\mathtt{M}}$. Thus, we finally obtain

$$\begin{aligned} J_{1a}(d_0) &= \pi_w(d_0)\left[2R_\nu(\delta^{\mathtt{M}},\alpha_h)+\kappa\right]-\frac{\kappa}{\rho(d_0)} \\ &\quad+\max_{\delta\in[0,\nu]}\bar{\pi}_w(d_0)\left[R_\nu(\delta,\alpha_\ell)+\pi_w(d_0+\delta)R_\nu(\delta^{\mathtt{M}},\alpha_h)+\bar{\pi}_w(d_0+\delta)R_\nu(\delta^{\mathtt{M}},\alpha_\ell)\right]-\frac{\kappa}{\rho(d_0+\delta)}\,. \\ &= \pi_w(d_0)\left[2R_\nu(\delta^{\mathtt{M}},\alpha_h)+\kappa\right]-\frac{\kappa}{\rho(d_0)}+\bar{\pi}_w(d_0)\tau\alpha_\ell \\ &\quad+\max_{\delta\in[0,\nu]}\bar{\pi}_w(d_0)\left[R_\nu(\delta,\alpha_\ell)+\pi_w(d_0+\delta)\tau(\alpha_h-\alpha_\ell)\right]-\frac{\kappa}{\rho(d_0+\delta)}\,. \end{aligned} \tag{19}$$

The second equality follows from the fact that $R_\nu(\delta^{\mathtt{M}},\alpha_i)=\tau\alpha_i$ for $i\in\{h,\ell\}$ (see Lemma 1). Here, the uniqueness of $\delta=\delta^*$ follows because the objective function $\bar{\pi}_w(d_0)\left[R_\nu(\delta,\alpha_\ell)+\pi_w(d_0+\delta)\tau(\alpha_h-\alpha_\ell)\right]-\kappa/\rho(d_0+\delta)$ is concave and continuous in $\delta$ and $[0,\nu]$ is a compact interval. Q.E.D.

*Proof of Proposition 2.* After removing incentive compatibility constraints and setting all payments to $\kappa$ in Problem (6), we use Lemma 1 and obtain the following optimization problem

$$\begin{aligned} J^{\mathtt{FB}}_{1a}(d_0)=\max_{\delta_{1h},\delta_{1\ell}\in[0,\nu]} & \pi_w(d_0)[R_\nu(\delta_{1h},\alpha_h)+R_\nu(\delta^{\mathtt{M}},\alpha_h)-\kappa]+\bar{\pi}_w(d_0)\Big[R_\nu(\delta_{1\ell},\alpha_\ell) \\ & +\pi_w(d_0+\delta_{1\ell})R_\nu(\delta^{\mathtt{M}},\alpha_h)+\bar{\pi}_w(d_0+\delta_{1\ell})R_\nu(\delta^{\mathtt{M}},\alpha_\ell)-2\kappa\Big]\,. \end{aligned} \tag{20}$$

Isolating $\delta_{1h}$ and using Lemma 1 again, we obtain $\delta_{1h} = \delta^{\mathtt{M}}$. The remaining optimization problem is

$$J_{1a}^{\mathtt{FB}}(d_0) = \pi_w(d_0)[2R_\nu(\delta^{\mathtt{M}}, \alpha_h) - \kappa] + \bar{\pi}_w(d_0)(\tau\alpha_\ell - 2\kappa) + \bar{\pi}_w(d_0)\left\{\max_{\delta\in[0,\nu]} R_\nu(\delta, \alpha_\ell) + \pi_w(d_0+\delta)\tau(\alpha_h - \alpha_\ell)\right\}. \tag{21}$$

Because $\pi_w$, $R_\nu$ are concave and $d_0 + \delta$ is a linear function of $\delta$, the objective function inside curly brackets is concave in $\delta$. This implies that the optimal solution $\delta = \delta^{\mathtt{FB}}$ is unique.

We next prove $\delta^{\mathtt{FB}} > \delta^{\mathtt{M}}$. Recall that $\delta^{\mathtt{M}}$ is the unique optimal solution to $\max_{\delta\in[0,\nu]} R_\nu(\delta, \alpha_\ell)$, $R_\nu(\delta, \alpha_\ell)$ is concave (see Lemma 1), and $\pi_w(d_0 + \delta)$ is increasing in $\delta$. The marginal revenue for $R_\nu(\delta, \alpha_\ell)$ at $\delta = \delta^{\mathtt{M}}$ is 0 while $\pi_w'(d_0 + \delta^{\mathtt{M}})$ is positive. Therefore, it follows that

$$\Big[\frac{\partial R_\nu(\delta, \alpha_\ell)}{\partial\delta} + \frac{\partial \pi_w(d_0+\delta)(\alpha_h - \alpha_\ell)\tau}{\partial\delta}\Big]|_{\delta=\delta^{\mathtt{M}}} > 0\,.$$

Note that the objective function in (21) is concave in $\delta_{1\ell}$ and increasing when evaluated at $\delta_{1\ell} = \delta^{\mathtt{M}}$. Therefore, it follows that $\delta^{\mathtt{FB}} > \delta^{\mathtt{M}}$. Q.E.D.

*Proof of Theorem 1.* Note that if $\rho(d)$ is constant, then the objective function in Proposition 1 and the one in Proposition 2 differ from each other by a constant. Therefore, their optimal solutions are the same, i.e., $\delta^{\mathtt{FB}} = \delta^*$.

Next, assume that $\rho(d)$ is decreasing in $d$. We prove this by contradiction. Assume that $\delta^* > \delta^{\mathtt{FB}}$. Fix $d_0$ and $\nu$. The following condition is satisfied by $\delta^{\mathtt{FB}}$ because $\delta^{\mathtt{FB}}$ is the unique optimal solution of $\max_{\delta\in[0,\nu]} R_\nu(\delta, \alpha_\ell) + \pi_w(d_0 + \delta)\tau(\alpha_h - \alpha_\ell)$ (see Proposition 2).

$$R_\nu(\delta^{\mathtt{FB}}, \alpha_\ell) + \pi_w(d_0 + \delta^{\mathtt{FB}})\tau(\alpha_h - \alpha_\ell) \geq R_\nu(\delta, \alpha_\ell) + \pi_w(d_0 + \delta)\tau(\alpha_h - \alpha_\ell)\,, \forall\delta \in [0, \nu] \tag{22}$$

Because the data impact $\rho(\cdot)$ is decreasing, the term $-\dfrac{\kappa}{\rho(d_0+\delta)}$ is an increasing function of $\delta$. Using the assumption of contradiction, we obtain

$$-\frac{\kappa}{\rho(d_0 + \delta^{\mathtt{FB}})} > -\frac{\kappa}{\rho(d_0 + \delta^*)} \tag{23}$$

Inequalities (22) and (23) imply that the objective function $\bar{\pi}_w(d_0)[R_\nu(\delta, \alpha_\ell) + \pi_w(d_0 + \delta)\tau(\alpha_h - \alpha_\ell)] - \dfrac{\kappa}{\rho(d_0+\delta)}$ evaluated at $\delta = \delta^{\mathtt{FB}}$ is strictly larger than the value obtained by evaluating the same at $\delta = \delta^*$. Therefore, the condition $\delta^* > \delta^{\mathtt{FB}}$ contradicts the fact that $\delta^*$ is the optimal solution (see Proposition 1). This implies that $\delta^* \leq \delta^{\mathtt{FB}}$. Furthermore, we know $\delta^{\mathtt{FB}} \neq \delta^*$ from the first-order conditions of the optimization problems in Propositions 1 and 2. Hence, we obtain $\delta^* < \delta^{\mathtt{FB}}$. The last item of the theorem follows the same steps by assuming $\delta^* < \delta^{\mathtt{FB}}$. Q.E.D.

*Proof of Theorem 2.* Before proving this result, we first provide a constructive lemma that shows that the social benefits are monotone in the firm's data collection decision at equilibrium.

LEMMA 3. *Social welfare $W^*$ and consumer surplus $C^*$ are nondecreasing in $\delta^*$. Furthermore, $W^* = W^{\mathtt{FB}}$ and $C^* = C^{\mathtt{FB}}$ for any $\delta^* = \delta^{\mathtt{FB}}$.*

(The proof of this lemma is relegated to the end of this proof.) From Theorem 1, we know $\delta^* = \delta^{\mathtt{FB}}$ ($\delta^* > \delta^{\mathtt{FB}}$ and $\delta^* < \delta^{\mathtt{FB}}$) when $\rho$ is constant (increasing and, respectively, decreasing). Furthermore, Lemma 3 implies that $W^* = W^{\mathtt{FB}}$ and $C^* = C^{\mathtt{FB}}$ ($W^* > W^{\mathtt{FB}}$ and $C^* < C^{\mathtt{FB}}$; and $W^* < W^{\mathtt{FB}}$ and $C^* > C^{\mathtt{FB}}$) when $\delta^* = \delta^{\mathtt{FB}}$ ($\delta^* > \delta^{\mathtt{FB}}$ and, respectively, $\delta^* < \delta^{\mathtt{FB}}$). Q.E.D.

*Proof of Lemma 3.* In a pricing period (e.g., Period 1b), the total consumer surplus for a given price $p$ and accuracy $\alpha$ is given by $\int_0^1 1_{\{\alpha v \geq p\}}(\alpha v - p) f(v)\mathrm{d}v = \alpha/2 - p + p^2/(2\alpha)$, where $1_{\{\cdot\}}$ is the indicator function. Similarly, the total social welfare for a given price $p$ and accuracy $\alpha$ is $\int_0^1 1_{\{\alpha v \geq p\}} \alpha v f(v)\mathrm{d}v = \alpha/2 - p^2/(2\alpha)$. In order to find the total expected social welfare and the customer surplus at equilibrium, we use the equilibrium prices and take expectation with respect to the induced probabilities of the amounts of collected data at equilibrium. The total expected social welfare and the customer surplus are in order.

$$\begin{aligned} W^* =& \pi_w(d_0)\alpha_h\left(1-(1-2\tau)^2\right) + \bar{\pi}_w(d_0)\frac{\alpha_\ell}{2}\left(1-\left(1-\frac{\delta^*}{\nu}\right)^2\right) \\ &+ \bar{\pi}_w(d_0)\left[\pi_w(d_0+\delta^*)\frac{\alpha_h}{2}\left(1-(1-2\tau)^2\right) + \bar{\pi}_w(d_0+\delta^*)\frac{\alpha_\ell}{2}\left(1-(1-2\tau)^2\right)\right] \\ C^* =& \pi_w(d_0)2\tau^2(2\alpha_h - \alpha_\ell) + 2\tau^2\alpha_\ell + \bar{\pi}_w(d_0)\frac{\alpha_\ell(\delta^*)^2}{2\nu^2} + \bar{\pi}_w(d_0)\pi_w(d_0+\delta^*)2\tau^2(\alpha_h - \alpha_\ell) \end{aligned}$$

The first-best benchmarks $W^{\mathtt{FB}}$ and $C^{\mathtt{FB}}$ are found by evaluating the terms above using $\delta^{\mathtt{FB}}$ instead of $\delta^*$. Since $\pi_w$ is increasing and $\delta^*, \delta^{\mathtt{FB}} \in [0, \nu]$, the monotonicity result follows. Q.E.D.

*Proof of Proposition 3.* We use the first-order condition for the maximization problem provided in Proposition 2 and the implicit function theorem to prove this result. The first-order condition is

$$\left[\frac{\partial R_\nu(\delta, \alpha_\ell)}{\partial \delta} + \frac{\partial \pi_w(d_0+\delta)\tau(\alpha_h - \alpha_\ell)}{\partial \delta}\right]\Big|_{\delta=\delta^{\mathtt{FB}}} = 0\,.$$

Note here that the left-hand side of the equality above is a continuously differentiable function of $\delta^{\mathtt{FB}}$ because $R_\nu(\delta, \alpha_\ell)$ and $\pi_w(d_0+\delta)$ are twice-differentiable with a continuous second-order derivative w.r.t. $\delta$; thus we can use the implicit function theorem. From the implicit function theorem, we know that $\frac{\partial \delta^{\mathtt{FB}}}{\partial \nu}$ has the same sign as $\frac{\partial^2 R_\nu(\delta, \alpha_\ell)}{\partial \nu \partial \delta}\Big|_{\delta=\delta^{\mathtt{FB}}} = \frac{\alpha_\ell}{\nu^2}\left(-1 + \frac{4\delta^{\mathtt{FB}}}{\nu}\right)$ because we consider monotonicity with respect to $\nu$. Recall that we know $\delta^{\mathtt{FB}} \geq \delta^{\mathtt{M}}$ from Proposition 2 and $\delta^{\mathtt{M}} = \nu\bar{F}(\phi^{-1}(0)) = \nu/2$ from Lemma 1. Therefore, $\delta^{\mathtt{FB}}$ is nondecreasing in $\nu$.

*Proof of Theorem 3.* We use the first-order condition for the maximization problem provided in Proposition 1 and the implicit function theorem to prove this result. The first-order condition is

$$\bar{\pi}_w(d_0)\left[\frac{\alpha_\ell}{\nu}\left(1 - \frac{2\delta^*}{\nu}\right) + \pi_w'(d_0+\delta^*)(\alpha_h - \alpha_\ell)\tau\right] + \frac{\kappa\rho'(d_0+\delta^*)}{[\rho(d_0+\delta^*)]^2} = 0\,. \tag{24}$$

Note that $1/\rho$ is twice-differentiable and with a continuous second-order derivative because $\pi_w$ and $\pi_s$ are so and $\pi_w(d)-\pi_s(d)>0$ for all $d\in[d_{\min},d_{\max}]$. Because the left-hand side of this equation is decreasing in $\delta^*$ (due to concavity), the implicit function theorem implies that the sign of $\dfrac{\partial\delta^*}{\partial\nu}$ is equal to $\dfrac{\alpha_\ell}{\nu^2}\left(-1+\dfrac{4\delta^*}{\nu}\right)$.

We know that if $\rho$ is increasing, $\delta^*\geq\delta^{\mathtt{FB}}$ (see Theorem 1), $\delta^{\mathtt{FB}}\geq\delta^{\mathtt{M}}$ (see Proposition 2) and $\delta^{\mathtt{M}}=\nu/2$ for the standard uniform distribution (see Lemma 1). Therefore, it follows that $\dfrac{\partial\delta^*}{\partial\nu}\geq 0$.

If $\rho$ is decreasing, it is possible that $\delta^*$ can take values smaller than $\nu/4$. If $\delta^*$ is smaller than $\nu/4$ for all $\nu\in(0,\bar\nu]$, it follows that $\delta^*$ is decreasing in $\nu$; therefore $\hat\nu$ takes the lowest possible $\nu$ value. If $\delta^*$ is larger than $\nu/4$ for all $\nu\in(0,\bar\nu]$, it follows that $\delta^*$ is increasing in $\nu$; therefore $\hat\nu=\bar\nu$. These two cases are degenerate in the sense that the monotonicity of $\delta^*$ does not change at an interior $\nu$.

On the other hand, if $\delta^*$ crosses $\nu/4$ at some point $\upsilon\in(0,\bar\nu]$, then $\delta^*$ must first be increasing, i.e., $\delta^*\geq\nu/4$ for all $\nu\in(0,\upsilon]$, and then be decreasing, i.e., $\delta^*\leq\nu/4$ for all $\nu\in[\upsilon,\bar\nu]$. This implies $\hat\nu=\upsilon$. Note that $\delta^*$ can cross $\nu/4$ at most once because $\delta^*$ is decreasing in $\nu$ after crossing but $\nu/4$ is increasing. We provide a sufficient condition for this case in Proposition EC.6 in Section EC.3 of the e-companion. Q.E.D.

*Proof of Theorem 4.* The firm's expected profit derived in Proposition 1 is

$$\begin{aligned}J_{1a}(d_0)=&\pi_w(d_0)\left[2R_\nu(\delta^{\mathtt{M}},\alpha_h)+\kappa\right]-\frac{\kappa}{\rho(d_0)}+\bar\pi_w(d_0)\tau\alpha_\ell\\&+\bar\pi_w(d_0)\left[R_\nu(\delta^*,\alpha_\ell)+\pi_w(d_0+\delta^*)\tau(\alpha_h-\alpha_\ell)\right]-\frac{\kappa}{\rho(d_0+\delta^*)}\,.\end{aligned}$$

We take the derivative of $J_{1a}(d_0)$ with respect to $\nu$. For simplicity, we denote by $\delta^*_\nu$ the derivative of $\delta^*$ with respect to $\nu$. Note that we consider interior optimal solutions $\delta^*$ that satisfy the first-order condition, so $\delta^*_\nu$ exists.

$$\begin{aligned}\frac{\partial J_{1a}(d_0)}{\partial\nu}=&\bar\pi_w(d_0)\left[\frac{\alpha_\ell}{\nu}\left(\delta^*_\nu-\frac{\delta^*}{\nu}\right)\left(1-\frac{2\delta^*}{\nu}\right)+\pi_w(d_0+\delta^*)\tau(\alpha_h-\alpha_\ell)\delta^*_\nu\right]+\frac{\kappa\rho'(d_0+\delta^*)\delta^*_\nu}{[\rho(d_0+\delta^*)]^2}\\=&\delta^*_\nu\left\{\bar\pi_w(d_0)\left[\frac{\alpha_\ell}{\nu}\left(1-\frac{2\delta^*}{\nu}\right)+\pi_w(d_0+\delta^*)\tau(\alpha_h-\alpha_\ell)\right]+\frac{\kappa\rho'(d_0+\delta^*)}{[\rho(d_0+\delta^*)]^2}\right\}\\&+\bar\pi_w(d_0)\frac{\alpha_\ell}{\nu}\left(\frac{\delta^*}{\nu}\right)\left(\frac{2\delta^*}{\nu}-1\right).\end{aligned}$$

From the first-order condition for the maximization problem in Proposition 1 and the implicit function theorem, we know that

$$\bar\pi_w(d_0)\left[\frac{\alpha_\ell}{\nu}\left(1-\frac{2\delta^*}{\nu}\right)+\pi'_w(d_0+\delta^*)(\alpha_h-\alpha_\ell)\tau\right]+\frac{\kappa\rho'(d_0+\delta^*)}{[\rho(d_0+\delta^*)]^2}=0\,. \tag{25}$$

Thus, it follows that $\dfrac{\partial J_{1a}(d_0)}{\partial\nu}=\bar\pi_w(d_0)\dfrac{\alpha_\ell}{\nu}\left(\dfrac{\delta^*}{\nu}\right)\left(\dfrac{2\delta^*}{\nu}-1\right)=\bar\pi_w(d_0)\dfrac{\alpha_\ell}{\nu}\left(\dfrac{\delta^*}{\nu}\right)\dfrac{2}{\nu}\left(\delta^*-\dfrac{\nu}{2}\right)$.

The sign of $\partial J_{1a}(d_0)/\partial\nu$ is determined by $\delta^* - \nu/2$. When $\rho$ is increasing, $\delta^* > \delta^{\mathtt{M}} = \nu/2$; thus $\partial J_{1a}(d_0)/\partial\nu$ is positive.

Recall that $\delta = q\nu$ by definition. Thus, we denote by $q^*$ the quantity corresponding to $\delta^*$ and focus on it to prove the second item of the theorem. We next provide a lemma whose proof is relegated to the end of the proof of this theorem.

Lemma 4. *When $\rho$ is decreasing, $q^*$ is decreasing in $\nu$ if $q^* < 1/2$.*

Note also that the sign of $\partial J_{1a}(d_0)/\partial\nu$ is determined by the sign of $(q^* - 1/2)$. When $\rho$ is decreasing, $\delta^*$ can be larger or smaller than $\delta^{\mathtt{M}} = \nu/2$; i.e., $q^*$ can be larger or smaller than $1/2$. We know that $\partial J_{1a}(d_0)/\partial\nu$ is positive as long as $q^* > 1/2$. If $q^*$ crosses $1/2$ at some $\tilde{\nu}$, then it follows that $\partial J_{1a}(d_0)/\partial\nu$ is negative for all $\nu \geq \tilde{\nu}$ because $q^*$ is decreasing if $q^* \leq 1/2$ when $\rho$ decreasing (see Lemma 4).

Note that the expected profit is decreasing in $\nu$ when $\rho$ is decreasing and $q^*$ is smaller than $1/2$. Thus, we next show that $q^*$ is decreasing in $d_0$ when $\rho$ is decreasing. To do so, we use the implicit function theorem with the first-order condition for the maximization problem in Proposition 1 in terms of $q^*$ given by

$$\frac{1}{2\alpha_\ell}\pi_w'(d_0 + q^*\nu)(\alpha_h - \alpha_\ell)\tau\nu + \frac{1}{2\bar{\pi}_w(d_0)\alpha_\ell}\frac{\kappa\rho'(d_0 + q^*\nu)\nu}{[\rho(d_0 + q^*\nu)]^2} = q^* - \frac{1}{2}\,. \tag{26}$$

The derivative of $q^*$ with respect to $d_0$ has the same sign with the following term:

$$\begin{aligned}&-\pi_w'(d_0)[\alpha_\ell(1-2q^*) + \pi_w'(d_0 + q^*\nu)\tau(\alpha_h - \alpha_\ell)\nu] + \bar{\pi}_w(d_0)\pi_w''(d_0 + q^*\nu)\tau(\alpha_h - \alpha_\ell)\nu \\ &+ \frac{\kappa\rho''(d_0 + q^*\nu)\nu}{[\rho(d_0 + q^*\nu)]^2} - \frac{2\kappa[\rho'(d_0 + q^*\nu)]\nu}{[\rho(d_0 + q^*\nu)]^3}\,.\end{aligned}$$

Replacing $q^*$ in the above expression using the left-hand side of (26), we obtain

$$\frac{\pi_w'(d_0)\kappa\rho'(d_0 + q^*\nu)\nu}{[\rho(d_0 + q^*\nu)]^2\bar{\pi}_w(d_0)} + \bar{\pi}_w(d_0)\pi_w''(d_0 + q^*\nu)\tau(\alpha_h - \alpha_\ell)\nu + \frac{\kappa\rho''(d_0 + q^*\nu)\nu}{[\rho(d_0 + q^*\nu)]^2} - \frac{2\kappa[\rho'(d_0 + q^*\nu)]\nu}{[\rho(d_0 + q^*\nu)]^3}\,.$$

Here, the expression above is negative because $\rho$ is decreasing, $1/\rho$ is convex; and $\pi_w$ is an increasing concave function. This concludes the proof. Q.E.D.

*Proof of Lemma 4.* Using (26) and the implicit function theorem, we know that the sign of the derivative of $q^*$ with respect to $\nu$ is the same as

$$\underbrace{\left[\bar{\pi}_w(d_0)\pi_w'(d_0 + q^*\nu)\tau(\alpha_h - \alpha_\ell) + \frac{\kappa\rho'(d_0 + q^*\nu)}{[\rho(d_0 + q^*\nu)]^2}\right]}_{(i)} \tag{27}$$

$$+ \underbrace{\left[\bar{\pi}_w(d_0)\pi_w''(d_0 + q^*\nu)\tau(\alpha_h - \alpha_\ell) + \frac{\kappa\rho''(d_0 + q^*\nu)}{[\rho(d_0 + q^*\nu)]^2} - \frac{2\kappa(\rho'(d_0 + q^*\nu))^2}{[\rho(d_0 + q^*\nu)]^3}\right]\nu q^*}_{(ii)}\,. \tag{28}$$

First note that $q^*$ can take values smaller than $1/2$ when $\rho$ is decreasing. If $q^* \leq 1/2$, the first-order condition (26) implies that $(i)$ is negative. We also know that $(ii)$ is always negative because $\pi_w$ is concave and $1/\rho$ is convex. Thus, $q^*$ starts to decrease after crossing $1/2$. Q.E.D.

# Electronic Companion

## EC.1 Interior Price Decisions of the Firm

In this section, we show that if the probability of high accuracy $\pi_w$ and the data impact $\rho$ do not increase or decrease too fast, the firm's optimal price (and hence the collected data size) take interior values. Let $\Pi'_{\max} \triangleq \max_{d\in[d_{\min},d_{\max}]} \pi'_w(d)$ and $\Pi'_{\min} \triangleq \min_{d\in[d_{\min},d_{\max}]} \pi'_w(d)$ and $D'_{\max} \triangleq \max_{d\in[d_{\min},d_{\max}]} \rho'(d)$ and $D'_{\min} \triangleq \min_{d\in[d_{\min},d_{\max}]} \rho'(d)$. Because $\pi_w$ is concave, in fact $\Pi'_{\max} = \pi'_w(d_{\min})$ and $\Pi'_{\min} = \pi'_w(d_{\max})$.

Proposition EC.1. *We have that*

- *If* $\tau(\alpha_h - \alpha_\ell)\Pi'_{\max} < 2\alpha_\ell/d_{\max}$, *then* $\delta^{\mathtt{FB}} \in (0,\nu)$ *for* $\nu \in (0,\nu_{\max}]$.
- *If* $\tau(\alpha_h - \alpha_\ell)\Pi'_{\max}\bar{\pi}_w(d_{\min}) + \kappa D'_{\max}/\underline{\Delta\pi}^2 < \bar{\pi}_w(d_{\max})2\alpha_\ell/d_{\max}$ *and* $\bar{\pi}_w(d_{\max})\tau(\alpha_h - \alpha_\ell)\Pi'_{\min} + \kappa D'_{\min} + \bar{\pi}_w(d_{\max})\alpha_\ell/\nu > 0$, *then* $\delta^* \in (0,\nu)$ *for* $\nu \in (0,\nu_{\max}]$.

*Proof of Proposition EC.1.* This result follows from the fact that the derivatives of the objective functions in the maximization problems in Proposition 2 and respectively Proposition 1 are positive at $\delta = 0$ and negative at $\delta = \nu$ when the conditions in the statement of the proposition are satisfied. (Recall that the objective functions in both problems are concave in $\delta$, see the proofs of Propositions 2 and 1.)

Note that Proposition 2 implies that $\delta^{\mathtt{FB}} \geq \delta^{\mathtt{M}}$ (and recall that $\delta^{\mathtt{M}} > 0$ from Lemma 1). Thus, we need to check the derivative for the first-best data size only at $\delta = \nu$. Using the definition of $R_\nu(\delta,\alpha_\ell)$, we obtain the derivative of the objective function in Proposition 2 as follows:

$$\alpha_\ell\left(\frac{1}{\nu} - \frac{2\delta}{\nu^2}\right) + \tau(\alpha_h - \alpha_\ell)\pi_w(d_0+\delta)\,. \tag{EC.1}$$

The condition in the first bullet point of the result implies that the term in (EC.1) is negative when evaluated at $\delta = \nu$. Hence, it follows that $\delta^{\mathtt{FB}} \in (0,\nu)$.

Next, we consider $\delta^*$. Similarly, the derivative of the objective function in Proposition 1 with respect to $\delta$ is given as follows:

$$\bar{\pi}_w(d_0)\left[\alpha_\ell\left(\frac{1}{\nu} - \frac{2\delta}{\nu^2}\right) + \tau(\alpha_h - \alpha_\ell)\pi'_w(d_0+\delta)\right] + \frac{\kappa\rho'(d_0+\delta)}{[\rho(d_0+\delta)]^2}\,. \tag{EC.2}$$

The first condition in the second bullet point guarantees that the term in (EC.2) evaluated at $\delta = \nu$ is negative for any $\nu \in (0,\nu_{\max}]$ because $\pi'_w(d_0+\nu) \leq \Pi'_{\max}$ and $\rho(d_0+\delta) \geq \underline{\Delta\pi}$. The second condition in the same bullet point guarantees that the term in (EC.2) is positive when evaluated at $\delta = 0$. Therefore, it follows that $\delta^* \in (0,\nu)$. Q.E.D.

Although the conditions in Proposition EC.1 are sufficient to guarantee that the optimal solutions $\delta^*$ and $\delta^{\mathsf{FB}}$ are interior, they are in fact loose because we do not assume any parametric form of functions $\pi_w$ and $\rho$. For given parametric forms of functions $\pi_w$ and $\rho$, these conditions can be improved. Alternatively, it is straightforward to check if the optimal solutions $\delta^*$ and $\delta^{\mathsf{FB}}$ are interior for any given set of parameters as in the numerical examples provided in the main body of our paper.

In case of the corner solution such that the $\delta^*$ takes the value of its upper bound $\nu$, the corresponding price of the product $p^*$ becomes 0 per Section 3. In this case, the firm distributes the product for free to collect the largest available data. We conclude this section by providing a necessary and sufficient condition for $\delta^* = \nu$.

Lemma EC.1. *The optimal data collection size $\delta^*$ equals to $\nu$ if and only if the following condition holds.*

$$\bar{\pi}_w(d_0)\tau(\alpha_h - \alpha_\ell)\pi'_w(d_0+\nu) + \frac{\kappa\rho'(d_0+\nu)}{[\rho(d_0+\nu)]^2} \geq \frac{\bar{\pi}_w(d_0)\alpha_\ell}{\nu} \tag{EC.3}$$

*Proof of Lemma EC.1.* The firm's optimal data collection size is found by maximizing the objective function in Proposition 1. When the inequality in (EC.3) holds, the objective function is nondecreasing for any $\delta \in [0,\nu]$ because it is concave. Thus, the firm selects the largest data collection size in this case.

## EC.2 Sufficient Conditions for nondegenerate $\tilde{\nu}$ and $\hat{\nu}$

Note that the threshold $\tilde{\nu}$ provided in Theorem 4 can take boundary values which makes the firm's profit entirely increasing. In this section, we provide a condition that guarantees that $\tilde{\nu}$ becomes nondegenerate. We focus on decreasing data impact $\rho$ because the firm's profit is increasing in $\nu$ otherwise.

Proposition EC.2. *Assume that $\rho$ is decreasing and define $a(\nu) \triangleq \frac{\pi'_w(d_0+\nu/2)\bar{\pi}_w(d_0)[\rho(d_0+\nu/2)]^2\tau}{-\rho'(d_0+\nu/2)}$. If $\lim_{\nu\to\infty} a(\nu) = 0$ then $\tilde{\nu}$ is finite.*

*Proof of Proposition EC.2.* We first prove that $\text{sign}(\partial J_{1a}(d_0)/\partial\nu) = \text{sign}(A(\nu))$ for any given $\nu$ where

$$A(\nu) \triangleq \pi'_w(d_0+\nu/2)\tau(\alpha_h - \alpha_\ell) + \frac{\kappa\rho'(d_0+\nu/2)}{\bar{\pi}_w(d_0)[\rho(d_0+\nu/2)]^2}.$$

In the proof of Theorem 4, we show that $\partial J_{1a}(d_0)/\partial\nu \geq 0$ if and only if $q^* \geq 1/2$ vice versa. Assume $A(\nu) \geq 0$, then the derivative of the objective function of the maximization problem in Proposition 1 evaluated at $q = 1/2$ is nonnegative which implies that $q^* \geq 1/2$ due to concavity. The reverse directions hold from the same relationship.

The statement in the proposition follows from the fact that $A(\nu) < 0$ if and only if $a(\nu) < \kappa/(\alpha_h - \alpha_\ell)$. Note that $\kappa/(\alpha_h - \alpha_\ell) > 0$, so $\lim_{\nu\to\infty} a(\nu) = 0$ implies that $a(\nu)$ becomes lower than

$\kappa/(\alpha_h - \alpha_\ell)$ for a sufficiently large $\nu$. For the same $\nu$, $A(\nu) < 0$. This implies that $\tilde{\nu}$ is finite. Q.E.D.

The condition on $a(\nu)$ provided in Proposition EC.2 guarantees that $\tilde{\nu} \in [0, \infty)$. The next corollary identifies a property on $\pi_w$ and $\pi_s$ that implies the limit condition in Proposition EC.2.

Proposition EC.3. *If $\pi_w(d) - \pi_s(d)$ is decreasing in $d$ and converges to 0 as $d$ approaches infinity, then $\lim_{\nu\to\infty} a(\nu) = 0$ and hence $\tilde{\nu} < \infty$.*

*Proof of Proposition EC.3.* Since $\pi_w(d) - \pi_s(d)$ is decreasing in $d$, $\rho(d)$ is also decreasing in $d$. We next check the limit of $a(\nu)$. To do so, we reformulate $a(\nu)$ as follows:

$$a(\nu) = \bar{\pi}_w(d_0)\tau \frac{\pi_w'(d_0+\nu/2)[\rho(d_0+\nu/2)]^2}{-\rho'(d_0+\nu/2)} = \bar{\pi}_w(d_0)\tau \frac{[\pi_w(d_0+\nu/2) - \pi_s(d_0+\nu/2)]^2}{\frac{\pi_s'(d_0+\nu/2)}{\pi_w'(d_0+\nu/2)}\pi_w(d_0+\nu/2) - \pi_s(d_0+\nu/2)}.$$

Note that $\pi_s'(d_0+\nu/2)/\pi_w'(d_0+\nu/2)$ is larger than 1 since $\pi_w(d) - \pi_s(d)$ is decreasing and $\pi_e(d)$ for $e \in \{s, w\}$ are increasing in $d$. Therefore, it follows that $a(\nu) < \bar{\pi}_w(d_0)\tau[\pi_w(d_0+\nu/2) - \pi_s(d_0+\nu/2)]$. Hence the sandwich theorem implies that $\lim_{\nu\to\infty} a(\nu) = 0$ because $\pi_w(d) - \pi_s(d)$ converges to 0 as $d$ approaches $\infty$ and $a(\nu) \geq 0$. Q.E.D.

The exponential function family ($\pi_e(d) = 1 - \exp(-z_e d)$) with parameter $z_e > 0$ for $e \in \{s, w\}$), and the power function family ($\pi_e(d) = 1 - y_e d^{-z_e}$ with parameters $y_e, z_e > 0$ for $e \in \{s, w\}$) are commonly used to estimate learning curves (see Table 1 in Viering and Loog 2021 for a list of references). The term $\pi_w(d) - \pi_s(d)$ converges to 0 as $d$ approaches infinity for these families. For the exponential family with parameters $z_e > 0$ for $e \in \{s, w\}$, the term $\pi_w(d) - \pi_s(d)$ is decreasing in $d$ over $[d_{\min} = \log(z_w/z_s)/(z_w - z_s), \infty)$. Note that $z_w > z_s$ in order to ensure $\pi_w(d) > \pi_s(d)$ for this family and hence $d_{\min} = \log(z_w/z_s)/(z_w - z_s) > 0$. For the power function family with parameters $y_e, z_e > 0$ for $e \in \{s, w\}$, the term $\pi_w(d) - \pi_s(d)$ is decreasing in $d$ over $[d_{\min} = \log(z_w y_w/z_s y_s)/(z_w - z_s), \infty)$. Also note that $z_w > z_s$ and $y_w > y_s$ in order to ensure that $\pi_w(d) > \pi_s(d)$ for this family and hence $d_{\min} = \log(z_w y_w/z_s y_s)/(z_w - z_s) > 0$. Thus, these families, satisfy the conditions in Proposition EC.3, which also implies that $\lim_{\nu\to\infty} a(\nu) = 0$ for $a(\nu)$ defined in Proposition EC.2.

Next, we provide a sufficient condition for $\hat{\nu}$ provided in Theorem 3 to be in $[0, \infty)$.

Proposition EC.4. *Assume that $\rho$ is decreasing and define $B(\nu) \triangleq \pi_w'(d_0+\nu/4)\tau(\alpha_h - \alpha_\ell)\nu + \frac{\kappa\rho'(d_0+\nu/4)\nu}{\bar{\pi}_w(d_0)[\rho(d_0+\nu/4)]^2}$. If there exists a finite $\epsilon$ such that $B(\epsilon) < -\alpha_\ell/2$, then $\hat{\nu} < \epsilon$.*

*Proof of Proposition EC.4.* In the proof of Theorem 3, we show that $\partial\delta^*/\partial\nu \geq 0$ if and only if $q^* \geq 1/4$ vice versa. Assume there exists $\epsilon$ such that $B(\epsilon) < -\alpha_\ell/2$, then the derivative of the objective function of the maximization problem in Proposition 1 evaluated at $q = 1/4$ is negative which implies that $q^* \leq 1/4$. Therefore, $q^*$ crosses $1/4$ at $\hat{\nu}$ and $\hat{\nu} < \epsilon$ due to concavity in $q$. Q.E.D.

## EC.3 Optimal Effort Choice of the Firm

In our setup, it may not be always optimal for the firm to induce the provider to work in both periods. To address this issue, we characterize an upper bound $\bar{\kappa}$ on the effort cost $\kappa$ depending on the problem parameters, and assume that $\kappa < \bar{\kappa}$. Under this assumption, the firm always finds it optimal to retrain the algorithm after observing a low accuracy $\alpha_\ell$ in the first period if $\rho$ is increasing. Furthermore, the condition $\kappa < \bar{\kappa}$ also guarantees that if $\rho$ is decreasing, the firm induces the provider to work in both periods for any $\nu \leq \bar{\nu}$ for some $\bar{\nu} > 0$, and $\tilde{\nu}$ provided in Theorem 4 is strictly lower than $\bar{\nu}$. The next proposition formally characterizes $\bar{\kappa}$.

Proposition EC.5. *There exists two positive thresholds $\bar{\kappa}$ and $\bar{\nu}$ such that the firm finds it optimal to induce provider to work in both periods if $\kappa < \bar{\kappa}$ and $\nu < \bar{\nu}$. Further, $\bar{\nu}$ equals $\infty$ when $\kappa < \bar{\kappa}$ and $\rho$ is increasing.*

*Proof of Proposition EC.5.* Let $\bar{\kappa} \triangleq \bar{\pi}_w(d_0)\tau(\alpha_h - \alpha_\ell)[\pi_w(d_0) - \pi_s(d_0)]$. Since the firm commits to the contract before the accuracy realizations, we compare the expected profit $J_{1a}(d_0)$ with the expected profit (denoted by $\tilde{J}_{1a}(d_0)$) collected when the firm hires the provider only for the product development. The expected profit $\tilde{J}_{1a}(d_0)$ for the no-retraining case is given by

$$\tilde{J}_{1a}(d_0) = \pi_w(d_0)2\alpha_h\tau + \bar{\pi}_w(d_0)2\alpha_\ell\tau - \frac{\kappa}{\rho(d_0)}. \tag{EC.4}$$

Because there is no retraining, in both periods the product is sold with the initial accuracy that can be $\alpha_h$ with probability $\pi_w(d_0)$ and $\alpha_\ell$ with the remaining probability. In each case, it is optimal for the firm to use the myopic price, therefore the expected revenue of the firm is given by $\pi_w(d_0)2\alpha_h\tau + \bar{\pi}_w(d_0)2\alpha_\ell$. The provider is hired only for the first period. In this case, the optimal contract for the provider is such that there is a payment of $\kappa/[\pi_w(d_0) - \pi_s(d_0)]$ in case of a high accuracy, and no payment otherwise. This causes an expected cost of $\kappa/\rho(d_0)$.

First assume that $\rho$ is increasing. The difference between this expected profits is

$$J_{1a}(d_0) - \tilde{J}_{1a}(d_0) = \pi_w(d_0)\kappa + \bar{\pi}_w(d_0)\left[R_\nu(\delta^*, \alpha_\ell) + \pi_w(d_0+\delta^*)\tau(\alpha_h - \alpha_\ell) - \frac{\kappa}{\bar{\pi}_w(d_0)\rho(d_0+\delta^*)} - \tau\alpha_\ell\right].$$

Note that the firm's profit is increasing in $\nu$ when $\rho$ is increasing. Therefore, we consider the difference term above when $\nu = 0$ and hence obtain the following lower bound.

$$J_{1a}(d_0) - \tilde{J}_{1a}(d_0) > \pi_w(d_0)\kappa + \bar{\pi}_w(d_0)\left[\pi_w(d_0)\tau(\alpha_h - \alpha_\ell) - \frac{\kappa}{\bar{\pi}(d_0)\rho(d_0)}\right] \tag{EC.5}$$

$$= \pi_w(d_0)\kappa + \bar{\pi}_w(d_0)\frac{\pi_w(d_0)}{\bar{\pi}_w(d_0)[\pi_w(d_0) - \pi_s(d_0)]}[\bar{\kappa} - \kappa] > 0 \tag{EC.6}$$

Since $\bar{\kappa} > \kappa$ all terms in the right-hand side of the above inequality are positive. Hence, the claim follows.

Next assume that $\rho$ is decreasing. In this case, the profit is increasing for $\nu \leq \tilde{\nu}$, and then decreasing $\nu \geq \tilde{\nu}$. As shown for the increasing $\rho$ case, the condition $\bar{\kappa} > \kappa$ ensures that the firm is better off retraining the algorithm when $\nu = 0$. Therefore, $J_{1a}(d_0)$ crosses $\tilde{J}_{1a}(d_0)$ at most one point for $\nu > 0$ because $J_{1a}(d_0)$ is unimodal and $\tilde{J}_{1a}(d_0)$ is constant in $\nu$. If exists, the crossing point is $\bar{\nu}$. Otherwise, $\bar{\nu}$ is $\infty$. Either case $\bar{\nu}$ is positive. Q.E.D.

Recall that we provide a condition in Section EC.2 of this e-companion to ensure $\tilde{\nu}$ is nondegenerate. The following proposition shows that the threshold $\bar{\nu}$ provided in Proposition EC.5 is strictly larger than $\tilde{\nu}$ when $\tilde{\kappa} > \kappa$ and $\lim_{\nu \to \infty} a(\nu) = 0$. Furthermore, this result also provides a sufficient condition for $\hat{\nu} < \bar{\nu}$.

Proposition EC.6. *We have the following*

- *If $\rho$ is decreasing, $\bar{\kappa} > \kappa$ and $\lim_{\nu \to \infty} a(\nu) = 0$, then $\tilde{\nu} < \bar{\nu}$.*
- *If $\rho$ is decreasing, $\bar{\kappa} > \kappa$ and there exists $B(\epsilon) < -\alpha_\ell/2$ such that $\epsilon < \bar{\nu}$, then $\hat{\nu} < \bar{\nu}$.*

*Proof of Proposition EC.6.* When $\rho$ is decreasing and $\kappa$ is lower than $\bar{\kappa}$, we know that $\bar{\nu}$ is positive. If $\bar{\nu}$ is infinite, then $\tilde{\nu} < \bar{\nu}$ because Proposition EC.2 implies $\tilde{\nu} < \infty$ when $\lim_{\nu \to \infty} a(\nu) = 0$. If $\bar{\nu}$ is finite, then we also know the point $\bar{\nu}$ where $J_{1a}(d_0)$ crosses $\tilde{J}_{1a}(d_0)$ is such that $\bar{\nu} > \tilde{\nu}$ because i) $J_{1a}(d_0)$ is unimodal with peak point $\tilde{\nu} \geq 0$, ii) $J_{1a}(d_0) > \tilde{J}_{1a}(d_0)$ at $\nu = 0$ when $\bar{\kappa} > \kappa$, and iii) $\tilde{J}_{1a}(d_0)$ is constant in $\nu$. The second part of the proposition directly follows from Proposition EC.4 because $\hat{\nu} < \epsilon$. Q.E.D.

## EC.4 Competitive Provider Market

In this section, we consider the setting where the firm can terminate the relation with the provider after observing a low accuracy in the first period, and contract with another provider in the second period. Recall that the firm does not need to retrain the algorithm after a high accuracy in the first period. Thus, firing after a low accuracy terminates the contract with the provider in the first period in any case. The firm then solves two separate contracting problems between which it collects data. To distinguish from the setting with a single provider, in this section we denote by $\theta$ data collection. The next proposition based on Laffont and Martimort (2009) characterizes the optimal payments of the firm.

Proposition EC.7. *In case of contracting with two providers in different periods, the optimal payments of the firm are such that $x^*_{1\ell} = 0$, $x^*_{2\ell} = 0$; and*

$$x^*_{1h} = \frac{\kappa}{\pi_w(d_0) - \pi_s(d_0)} \text{ and } x^*_{2h} = \frac{\kappa}{\pi_w(d_0 + \theta^*) - \pi_s(d_0 + \theta^*)}, \tag{EC.7}$$

*where $\theta^*$ is the unique optimal solution of $\max_{\theta \in [0,\nu]} R_\nu(\theta, \alpha_\ell) - \pi_w(d_0 + \theta)\tau(\alpha_h - \alpha_\ell) - \kappa/\rho(d_0 + \theta)$.*

*Proof of Proposition EC.7.* The optimal payments of the contracting problem with the first provider is characterized in Proposition 4.2 in Laffont and Martimort (2009). For any data collection $\theta$ after a low accuracy in the first period, the firm faces the same contracting problem in the second period but this time with a different accuracy probabilities: $\pi_w(d_0+\theta)$ and $\pi_s(d_0+\theta)$. Therefore, the same proposition in Laffont and Martimort (2009) characterizes the optimal payments in the second period for any given $\theta$.

At the end of the first period, the firm maximizes the next periods profit by choosing the optimal data collection size by anticipating the optimal payments when contracting with a new provider in case of obtaining low accuracy in the first period. Otherwise, goes with the myopic price for high accuracy. The firm's optimization problem after the low accuracy is as follows.

$$\max_{\theta\in[0,\nu]} \underbrace{R_\nu(\theta,\alpha_\ell)}_{\text{Current Revenue}} + \underbrace{\pi_w(d_0+\theta)\tau(\alpha_h-\alpha_\ell)}_{\text{Continuation Revenue}} - \underbrace{\kappa/\rho(d_0+\theta)}_{\text{Expected Agency Cost}} . \tag{EC.8}$$

Because all terms in the above optimization problem are concave functions of $\theta$ and $[0,\nu]$ is a compact set, there exists a unique $\theta^*$. Q.E.D.

As in the case of the single provider we restrict attention to $\theta^*\in(0,\nu)$ and the induced $q^*=\theta^*/\nu\in(0,1)$ for the following results.

Proposition EC.8. *When $\rho$ is decreasing, $q^*$ is decreasing in $d_0$.*

*Proof of Proposition EC.8.* As in the proof of Theorem 4, we use the first-order condition for (EC.8) and the implicit function theorem. The derivative of $q^*$ with respect to $d_0$ has the same sign with the following term:

$$\pi_w''(d_0+\theta^*)\nu(\alpha_h-\alpha_\ell)\tau + \frac{\kappa\rho''(d_0+\theta^*)\nu}{[\rho(d_0+\theta^*)]^2} - \frac{2\kappa[\rho'(d_0+\theta^*)]^2\nu}{[\rho(d_0+\theta^*)]^3} \tag{EC.9}$$

Because $\pi_w$ is concave and $1/\rho$ is convex it follows that the term above is negative. Q.E.D.

Note that the objective function in the optimization problem in (EC.8) differs from the one in Proposition 1 by a scaling term $\bar{\pi}_w(d_0)$ for the first two terms: current period revenue and continuation revenue. Therefore, Theorems 1 and 3 can be repeated verbatim using (EC.8). Thus, Theorem 2 is extended to this setting as well. The firm's expected profit in this setting is differently from Theorem 4 given by

$$\begin{aligned}&\pi_w(d_0)\left[2R_\nu(\delta^{\mathtt{M}},\alpha_h)+\kappa\right] - \frac{\kappa}{\rho(d_0)} + \bar{\pi}_w(d_0)\tau\alpha_\ell \\ &+\bar{\pi}_w(d_0)\left[R_\nu(\theta^*,\alpha_\ell)+\pi_w(d_0+\theta^*)\tau(\alpha_h-\alpha_\ell) - \frac{\kappa}{\rho(d_0+\theta^*)}\right].\end{aligned} \tag{EC.10}$$

Here, the difference is that the firm incurs the expected agency cost in the second period only if the first period yields low accuracy. Nevertheless the same steps in the proof of Theorem 4 (including

Lemma 4) can be followed verbatim because $\theta^*$ is the interior optimal solution to (EC.8) as $\delta^*$ is the interior optimal solution to the maximization problem in Proposition 1. To show Theorem 4 is also extended to this setting, we need to prove that $q^*$ is decreasing in $d_0$ when $\rho$ is decreasing in this setting, too. Proposition EC.8 does this.

The effect of incentive issues have further implications for these societal benefits. Specifically, our setup considers a unique provider, with which the firm contracts in both period. It is not clear, however, how the previous societal benefits change when the firm can contract with different providers in different periods. The next result shows that a more competitive market of providers may actually hurt the social benefits.

Proposition EC.9. *Let $\tilde{W}^*$ and $\tilde{C}^*$ be the social welfare and, respectively the customer surplus in a setting where the firm changes providers in each period.*

1. *If $\rho$ is increasing, $W^* > \tilde{W}^*$ and $C^* < \tilde{C}^*$.*
2. *If $\rho$ is decreasing, $W^* < \tilde{W}^*$ and $C^* > \tilde{C}^*$.*

Thus, a more competitive provider market hurts societal benefits more if the data impact increases. This is because the firm suffers more from the moral hazard of the provider in the absence of competition. Therefore, the firm collects more data when the data impact is increasing to mitigate incentive issues compared to the competitive provider market.

*Proof of Proposition EC.9.* This proposition stems from the fact that $\theta^*$ is larger than $\delta^*$ when $\rho$ is decreasing and vice versa. Specifically, we show in Lemma 3 that collecting more data increases both the social welfare and the customer surplus. Thus, to compare those societal benefits we can directly compare the optimal data collection sizes in two settings using the following lemma.

Lemma EC.2. *The optimal data collections $\delta^*$ and $\theta^*$ are such that*

- *$\theta^* < \delta^*$ if $\rho$ is increasing, and*
- *$\theta^* > \delta^*$ if $\rho$ is decreasing.*

Lemma 3 together with Lemma EC.2 imply the result. Q.E.D.

*Proof of Lemma EC.2.* The optimal data collection $\delta^*$ and $\theta^*$ satisfy the following first-order conditions.

$$\alpha_\ell\left(\frac{1}{\nu}-\frac{2\delta^*}{\nu^2}\right)+\tau(\alpha_h-\alpha_\ell)\pi_w'(d_0+\delta^*)+\frac{\kappa\rho'(d_0+\delta^*)}{\bar{\pi}_w(d_0)[\rho(d_0+\delta^*)]^2}=0$$

$$\alpha_\ell\left(\frac{1}{\nu}-\frac{2\theta^*}{\nu^2}\right)+\tau(\alpha_h-\alpha_\ell)\pi_w'(d_0+\theta^*)+\frac{\kappa\rho'(d_0+\theta^*)}{[\rho(d_0+\theta^*)]^2}=0$$

If $\rho$ is increasing, then $\alpha_\ell\left(\frac{1}{\nu}-\frac{2\theta^*}{\nu^2}\right)+\tau(\alpha_h-\alpha_\ell)\pi_w'(d_0+\theta^*)+\frac{\kappa\rho'(d_0+\theta^*)}{\bar{\pi}_w(d_0)[\rho(d_0+\theta^*)]^2}>0$ because $\bar{\pi}_w(d_0)<1$. Hence, $\theta^*<\delta^*$ because both objective functions are concave, hence their derivatives are decreasing. This implies the firm collects more data if the provider cannot be fired after a low accuracy in the

first period and the data impact is increasing. In the case of decreasing $\rho$, $\alpha_\ell\left(\frac{1}{\nu}-\frac{2\theta^*}{\nu^2}\right)+\tau(\alpha_h-\alpha_\ell)\pi_w'(d_0+\theta^*)+\frac{\kappa\rho'(d_0+\theta^*)}{\bar{\pi}_w(d_0)[\rho(d_0+\theta^*)]^2}<0$ so $\theta^*>\delta^*$. Q.E.D.

Finally, note that comparing (EC.10) with $J_{1a}(d_0)$ provided in the proof of Proposition 1 reveals that the firm always prefers the competitive provider market because the agency cost is lower.

## EC.5 Short-term Contracting

In this section, we consider a setting where the firm has no commitment power across periods when contracting with the provider. In order to differentiate the data collection sizes in this section from $\delta$ in our paper, we denote by $\gamma$ data collection sizes. Furthermore, we use the hat symbol $\hat{\cdot}$ over the value functions for this setting.

Since the firm has no commitment power, we formulate the firm's problem using backward induction starting from the last period. Denote then by $\hat{J}_{2b}(\alpha_2)$ the firm's optimal expected profit in Period 2b given accuracy $\alpha_2$ such that

$$\hat{J}_{2b}(\alpha_2)=\max_{\delta_2\in[0,\nu]} R_\nu(\delta_2,\alpha_2) \tag{EC.11}$$

The firm chooses the amount of collected data (or equivalently the price) so as to maximize the expected revenue in the current period as in (2).

We denote by $\hat{J}_{2a}(\alpha_1,d_1)$ the firm's optimal expected profit in Period 2a, given accuracy $\alpha_1$ and data size $d_1$, such that

$$\hat{J}_{2a}(\alpha_\ell,d_1) = \max_{x_{2h},x_{2\ell}\geq 0} \pi_w(d_1)[\hat{J}_{2b}(\alpha_h)-x_{2h}]+\bar{\pi}_w(d_1)[\hat{J}_{2b}(\alpha_\ell)-x_{2\ell}] \tag{EC.12}$$

$$\text{s.t.}$$

$$\pi_w(d_1)x_{2h}+\bar{\pi}_w(d_1)x_{2\ell}-\kappa\geq \pi_s(d_1)x_{2h}+\bar{\pi}_s(d_1)x_{2\ell} \tag{EC.13}$$

$$\hat{J}_{2a}(\alpha_h,d_1) = \hat{J}_{2b}(\alpha_h). \tag{EC.14}$$

When $\alpha_1=\alpha_\ell$, the firm needs to set payments such that the provider has enough incentives to exert effort, as formalized by incentive constraint (EC.13). These payments are then deduced from the firm's expected revenues in Period 2a. Here, data size $d_1$ affects the chance of improving accuracy in the next period via probabilities $\pi_e(\cdot)$, $e\in\{w,s\}$. When $\alpha_1=\alpha_h$, recall that the firm does not need nor pay the provider. We thus refer to $x_{2h}^*$ and $x_{2\ell}^*$ as the optimal payments solving Problem (EC.12).

Moving to the first period, we denote by $\hat{J}_{1b}(\alpha_1,d_0)$ the firm's optimal expected profit in Period 1b given accuracy $\alpha_1$ and data size $d_0$, such that

$$\hat{J}_{1b}(\alpha_1,d_0)=\max_{\gamma_1\in[0,\nu]} R_\nu(\gamma_1,\alpha_1)+\hat{J}_{2a}\big(\alpha_1,d_0+\gamma_1\big) \text{ for } \alpha_1\in\{\alpha_h,\alpha_\ell\}. \tag{EC.15}$$

In contrast to Problem (EC.11), the firm needs to balance the revenues in the current period with the expected profit in the next one when the algorithm is of low accuracy. When $\alpha_1 = \alpha_\ell$, the choice of data $\gamma_1$ (or equivalently price) affects current revenues directly and future ones indirectly by increasing the dataset size to $d_0 + \gamma_1$. We refer to $\gamma^*$ as the optimal solution of Problem (EC.15) for $\alpha_1 = \alpha_\ell$. When $\alpha_1 = \alpha_h$, no retraining is required and the optimal price is equal to the myopic price.

Given initial data size $d_0$, we denote by $\hat{J}_{1a}(d_0)$ the optimal total expected profit in Period 1a, such that

$$\hat{J}_{1a}(d_0) = \max_{x_{1h}, x_{1\ell} \geq 0} \pi_w(d_0)[\hat{J}_{1b}(\alpha_h, d_0) - x_{1h}] + \bar{\pi}_w(d_0)[\hat{J}_{1b}(\alpha_\ell, d_0) - x_{1\ell}] \tag{EC.16}$$

$$\text{s.t.}$$

$$\pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)\left[x_{1\ell} + \hat{J}_p(d_0)\right] - \kappa \geq \pi_s(d_0)x_{1h} + \bar{\pi}_s(d_0)\left[x_{1\ell} + \hat{J}_p(d_0)\right] \tag{EC.17}$$

$$\hat{J}_p(d_0) = \pi_w\left(d_0 + \delta^*\right)x^*_{2h} + \bar{\pi}_w\left(d_0 + \delta^*\right)x^*_{2\ell} - \kappa. \tag{EC.18}$$

The optimization problem of $\hat{J}_{2a}$ in (EC.12) corresponds to minimizing the expected payment while motivating the provider to exert effort. The following proposition (based on Laffont and Martimort 2009, Proposition 4.2) provides the corresponding optimal payments.

Proposition EC.10. *Given dataset size $d_1$, the unique optimal payments for Problem* (EC.12) *are $x^*_{2\ell} = 0$ and $x^*_{2h} = \kappa/[\pi_w(d_1) - \pi_s(d_1)]$. The firm's optimal expected profit is then*

$$\hat{J}_{2a}(\alpha_\ell, d_1) = \pi_w(d_1)\tau\alpha_h + \bar{\pi}_w(d_1)\tau\alpha_\ell - \frac{\kappa}{\rho(d_1)} \text{ and } \hat{J}_{2a}(\alpha_h, d_1) = \tau\alpha_h\,. \tag{EC.19}$$

*where $\kappa/\rho(d_1)$ is the expected payment to the provider.*

*Proof of Proposition EC.10.* Following Proposition 4.2 in Laffont and Martimort (2009, p. 157), we obtain the optimal payments for Problem (EC.12) $x^*_{2\ell} = 0$ and $x^*_{2h} = \kappa/[\pi_w(d_1) - \pi_s(d_1)]$. Following Lemma 1, we have $J_{2b}(\alpha_2, d_1) = \alpha_2\tau$. Combining these, we get

$$\hat{J}_{2a}(\alpha_\ell, d_1) = \pi_w(d_1)\tau\alpha_h + \bar{\pi}_w(d_1)\tau\alpha_\ell - \frac{\kappa}{\rho(d_1)}\,.$$

If $\alpha_1 = \alpha_h$, there are no payments so $\hat{J}_{2a}(\alpha_h, d_1) = \hat{J}_{2b}(\alpha_h) = \tau\alpha_h$. Q.E.D.

The optimal expected continuation profit $\hat{J}_{1b}(\alpha_1, d_0)$ in Problem (EC.15) is then obtained by using expected profit $\hat{J}_{2a}(\alpha_1, d_0 + \gamma)$ from (EC.19) in Proposition EC.10. We show next that the optimal data size solving Problem (EC.15) and hence the corresponding optimal price are unique

Proposition EC.11. *The optimal solution to Problem* (EC.15) *is unique and equal to $\delta^{\mathtt{M}}$ if $\alpha_1 = \alpha_h$.*

*Proof of Proposition EC.11.* Using the expected profit $\hat{J}_{2a}$ derived in (EC.19), we write Problem (EC.15) as follows.

$$\max_{\gamma_1 \in [0,\nu]} R_\nu(\gamma_1, \alpha_1) + 1\{\alpha_1 = \alpha_\ell\} \left[ \pi_w(d_0 + \gamma_1)\alpha_h \tau + \bar{\pi}_w(d_0 + \gamma_1)\alpha_\ell \tau - \frac{\kappa}{\rho(d_0 + \gamma_1)} \right] \tag{EC.20}$$

If $\alpha_1 = \alpha_h$, it is straightforward to see from (EC.20) that the unique optimal solution to Problem (EC.15) is $\delta^{\mathtt{M}}$. Otherwise, the optimal solution $\gamma^*$ solves the following problem

$$\max_{\gamma_1 \in [0,\nu]} R_\nu(\gamma_1, \alpha_\ell) + \pi_w(d_0 + \gamma_1)(\alpha_h - \alpha_\ell)\tau - \frac{\kappa}{\rho(d_0 + \gamma_1)} + \alpha_\ell \tau \tag{EC.21}$$

Because $\pi_w$, $R_\nu$ and $-1/\rho$ are concave, and $d_0 + \gamma_1$ is a linear function of $\gamma_1$, the objective function in (EC.21) is concave in $\gamma_1$. This implies that the optimal solution $\gamma^*$ is unique. Q.E.D.

Finally, the next result characterizes the optimal payments in Period 1a.

Proposition EC.12. *Given initial data size $d_0$, the unique optimal payments for Problem* (6) *are*

$$x^*_{1\ell} = 0 \text{ and } x^*_{1h} = \frac{\kappa}{\pi_w(d_0) - \pi_s(d_0)} + \left( \frac{\kappa}{\rho(d_0 + \gamma^*)} - \kappa \right).$$

*Proof of Proposition EC.12.* In Period 1a, the optimization problem of the firm is

$$\hat{J}_{1a}(d_0) = \pi_w(d_0)\hat{J}_{1b}(\alpha_h, d_0) + \bar{\pi}_w(d_0)\hat{J}_{1b}(\alpha_\ell, d_0) - \min_{x_{1h}, x_{1\ell} \geq 0} \pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)x_{1\ell}$$
$$\text{st. (EC.17), (EC.18)}$$

Using the optimal payments $x^*_{2h}$ and $x^*_{2\ell}$ derived in Proposition EC.10, we evaluate the expected continuation profit $\hat{J}_p(d_0)$ of the provider, and $\hat{J}_p(d_0) = \frac{\kappa}{\rho(d_0+\gamma^*)} - \kappa$. Characterizing $\hat{J}_p(d_0)$, we reduce this problem to a standard principal-agent model where the cost of effort is $\kappa + \left[\pi_w(d_0) - \pi_s(d_0)\right]\hat{J}_p(d_0)$. Therefore, Proposition 4.2 in Laffont and Martimort (2009, p. 157) implies that the optimal solution of the optimization problem in $\hat{J}_{1b}(d_0)$ is $x^*_{1\ell} = 0$ and $x^*_{1h} = \dfrac{\kappa}{\pi_w(d_0) - \pi_s(d_0)} + \dfrac{\kappa}{\rho(d_0 + \gamma^*)} - \kappa$. Q.E.D.

When focusing on $\gamma^* \in (0, \nu)$, we have that the objective function in the optimization problem in (EC.21) differs from the one in Proposition 1 by a scaling term $\bar{\pi}_w(d_0)$ for the first two terms: current period revenue and continuation revenue. Therefore, Theorems 1 and 3 can be repeated verbatim using (EC.21).

Proposition EC.13. *We have*

1. *If $\rho(\cdot)$ is strictly increasing, $\hat{J}_{1b}(\alpha_\ell, d_0)$ is strictly increasing in $\nu$.*
2. *If $\rho(\cdot)$ is strictly decreasing, a unique threshold $\nu_{st}$ exists such that $\hat{J}_{1b}(\alpha_\ell, d_0)$ is strictly increasing in $\nu$ if $\nu \leq \nu_{st}$ and is strictly decreasing otherwise. Further, threshold $d_{st} > 0$ exists such that $\hat{J}_{1b}(\alpha_\ell, d_0)$ is strictly decreasing (i.e., $\nu_{st} = 0$) if $d_0 \geq d_{st}$.*

*Proof of Proposition EC.13.* Note that $\hat{J}_{1b}(\alpha_\ell, d_0)$ is given in (EC.21). Using this and the implicit function theorem, we next derive $\partial \hat{J}_{1b}(\alpha_\ell, d_0)/\partial \nu$.

$$\frac{\partial \hat{J}_{1b}(\alpha_\ell, d_0)}{\partial \nu} = \frac{\alpha_\ell \gamma^*}{\nu^2}\left(\frac{2\gamma^*}{\nu} - 1\right).$$

Since Theorem 1 holds for this case as discussed above, it follows that $\gamma^* > \delta^{\mathtt{M}} = \nu/2$ when $\rho$ is increasing. This proves the first item. To proved the second item, we can follow the same steps in the proof of Theorem 4 but by using use Proposition EC.8. Q.E.D.

## EC.6 Withholding Data

*Proof of Theorem 5.* Differently from Problem (6), we allow for the firm to withhold a portion $\lambda \in [0,1]$ of the collected data $\delta_{1\ell}$ from the provider for the second period after a low accuracy in the first period. Note that the firm has no incentive to withhold data when $\rho$ is increasing because more data boosts the positive externalities of the AI Flywheel effect and mitigates the incentive issues. On the other hand, when the data impact $\rho$ is decreasing more data exacerbates the incentive issues hence the firm may want to withhold a portion of the data to compensate this effect. In this setup, the problem of the firm is as follows.

$$\max_{\substack{x_{1h}, x_{1\ell} \geq 0 \\ x_{2h}, x_{2\ell} \geq 0 \\ \delta_{1h}, \delta_{1\ell} \in [0,\nu] \\ \lambda \in [0,1]}} \pi_w(d_0)[R_\nu(\delta_{1h}, \alpha_h) + J_{2a}(\alpha_h, d_0 + \delta_{1h}) - x_{1h}] \tag{EC.22}$$

$$+\bar{\pi}_w(d_0)[R_\nu(\delta_{1\ell}, \alpha_\ell) + J_{2a}(\alpha_\ell, d_0 + \lambda\delta_{1\ell}) - x_{1\ell}]$$

s.t.

$$\pi_w(d_0)x_{1h} + \bar{\pi}_w(d_0)\left[x_{1\ell} + J_p(d_0)\right] - \kappa \geq$$

$$\pi_s(d_0)x_{1h} + \bar{\pi}_s(d_0)\left[x_{1\ell} + J_p(d_0)\right] \tag{EC.23}$$

$$J_p(d_0) = \pi_w\left(d_0 + \lambda\delta_{1\ell}\right)x_{2h} + \bar{\pi}_w\left(d_0 + \lambda\delta_{1\ell}\right)x_{2\ell} - \kappa. \tag{EC.24}$$

$$\pi_w(d_0 + \lambda\delta_{1\ell})x_{2h} + \bar{\pi}_w(d_0 + \lambda\delta_{1\ell})x_{2\ell} - \kappa \geq \tag{EC.25}$$

$$\pi_s(d_0 + \lambda\delta_{1\ell})x_{2h} + \bar{\pi}_s(d_0 + \lambda\delta_{1\ell})x_{2\ell}$$

For a fixed $\lambda\delta_{1\ell}$, we can use Lemmas 1 and 2 to derive the optimal payments. Hence, the firm's problem is reduced to choosing the optimal data collection size $\delta_{1\ell}$ and the portion $\lambda$.

$$\max_{\delta \in [0,\nu], \lambda \in [0,1]} \bar{\pi}_w(d_0)[R_\nu(\delta, \alpha_\ell) + \pi_w(d_0 + \lambda\delta)(\alpha_h - \alpha_\ell)\tau] - \frac{\kappa}{\rho(d_0 + \lambda\delta)} \tag{EC.26}$$

In this problem, the firm's pricing revenue $R_\nu(\delta, \alpha_\ell)$ does not involve $\lambda$ because the firm first determines the price of the product and collects revenue, then decides what portion ($\lambda$) of the collected data $\delta$ to share with the provider to improve the accuracy of the algorithm.

Using this expression, the expected profit of the firm can be written as follows:

$$\begin{aligned}&\pi_w(d_0)[2R_\nu(\delta^{\mathtt{M}},\alpha_h)+\kappa]-\frac{\kappa}{\rho(d_0)}+\bar{\pi}_w(d_0)\tau\alpha_\ell\\&+\max_{\delta\in[0,\nu],\lambda\in[0,1]}\bar{\pi}_w(d_0)[R_\nu(\delta,\alpha_\ell)+\pi_w(d_0+\lambda\delta)(\alpha_h-\alpha_\ell)\tau]-\frac{\kappa}{\rho(d_0+\lambda\delta)}\,.\end{aligned}\tag{EC.27}$$

In the expected profit of the firm, $R_\nu(\delta^{\mathtt{M}},\alpha_h)=\tau\alpha_h$ by Lemma 1. Therefore, the term $\nu$ appears only in the maximization problem. We next show that the optimal value of this maximization problem first increases in $\nu$ until $\nu<\nu_w$ for some $\nu_w$, then remains constant in $\nu$. By making a change of variables $\delta=q\nu$ and $\lambda\delta=\eta$, we can rewrite (EC.26) as follows.

$$\max_{q,\eta}\bar{\pi}_w(d_0)[\alpha_\ell q(1-q)+\pi_w(d_0+\eta)(\alpha_h-\alpha_\ell)\tau]-\frac{\kappa}{\rho(d_0+\eta)}\tag{EC.28}$$

$$\text{st. } 0\le q\le 1\tag{EC.29}$$

$$0\le\eta\le q\nu\tag{EC.30}$$

First note that, the term $\nu$ does not appear in the objective function but exists in the upper bound of the last constraint. Therefore, the optimal value is nondecreasing in $\nu$ because increasing $\nu$ enlarges the feasible region without affecting the objective function.

Assume there exists $\tilde{\eta}$ be such that $\bar{\pi}_w(d_0)\pi_w'(d_0+\tilde{\eta})(\alpha_h-\alpha_\ell)\tau=-\kappa\rho'(d_0+\tilde{\eta})/[\rho(d_0+\tilde{\eta})]^2$. If $\tilde{\eta}\le\nu/2$, $q^{\mathtt{M}}$ and $\eta^*=\tilde{\eta}$ is the optimal solution for (EC.28). In fact, it is the optimal solution of the problem obtained by relaxing the last constraint $\eta\le q\nu$. Observe that $\tilde{\eta}$ does not depend on $\nu$ because it is induced by functions $\pi_w$ and $\rho$. Therefore, increasing $\nu$ does not affect $\eta^*=\tilde{\eta}$ and $q^*=q^{\mathtt{M}}$, and the optimal value of (EC.26) hence the firm's expected profit is constant in $\nu$. Thus, one can define $\nu_w=2\tilde{\eta}$.

Next, we show that the expected profit is increasing if $\nu<\nu_w$. This can be shown that the last constraint in (EC.28) is binding when $2\tilde{\eta}>\nu$. We prove this result by contradiction. Assume that $2\tilde{\eta}>\nu$ and the last constraint is not binding at the optimal solution, i.e., $\eta^*-q^*\nu<0$.

Let $\mu$ be the Lagrange multiplier corresponding to the constraint $\eta-q\nu\le 0$. Thus, we obtain the following Karush-Kuhn-Tucker (KKT) conditions.

$$q^*=\frac{\alpha_\ell\bar{\pi}_w(d_0)+\mu\nu}{2\alpha_\ell\bar{\pi}_w(d_0)}$$

$$\bar{\pi}_w(d_0)\pi_w'(d_0+\eta^*)(\alpha_h-\alpha_\ell)\tau+\frac{\kappa\rho'(d_0+\eta^*)}{[\rho(d_0+\eta^*)]^2}=\mu$$

$$\mu(\eta^*-q^*\nu)=0$$

Since $\pi_w$ is concave and $1/\rho$ is convex, derivative of $\bar{\pi}(d_0)\pi_w(d_0+\eta)(\alpha_h-\alpha_\ell)\tau-\kappa/\rho(d_0+\eta)$ with respect to $\eta$ is decreasing. Since $\eta^*-q^*\nu<0$, then $\mu$ has to be 0 by complementary slackness condition $\mu(\eta^*-q^*\nu)=0$. This ($\mu=0$) also implies the optimal solution is $q^*=q^{\mathtt{M}}$ and $\eta^*=\tilde{\eta}$ which

also implies $2\tilde{\eta} < \nu$. This contradicts with the initial assumption $2\tilde{\eta} > \nu$. Therefore, $\eta^* = q^*\nu$ and increasing $\nu$ increases the expected profit when $2\tilde{\eta} > \nu$.

Finally, if $\bar{\pi}_w(d_0)\pi'_w(d_0+\tilde{\eta})(\alpha_h-\alpha_\ell)\tau + \kappa\rho'(d_0+\tilde{\eta})/[\rho(d_0+\tilde{\eta})]^2 > 0$ for all $\tilde{\eta}$, then we get $\mu > 0$ and hence $\nu_w = \nu_{\max}$. The case $\bar{\pi}_w(d_0)\pi'_w(d_0+\tilde{\eta})(\alpha_h-\alpha_\ell)\tau + \kappa\rho'(d_0+\tilde{\eta})/[\rho(d_0+\tilde{\eta})]^2 < 0$ for all $\tilde{\eta}$ implies that $\eta^* = 0$ and hence $\nu_w = 0$. Note that the term $\bar{\pi}_w(d_0)\pi'_w(d_0+\tilde{\eta})(\alpha_h-\alpha_\ell)\tau + \kappa\rho'(d_0+\tilde{\eta})/[\rho(d_0+\tilde{\eta})]^2$ is decreasing in $d_0$. Therefore, as $d_0$ is sufficiently large, i.e., $d_0 \geq d_w$ for some positive $d_w$, then $\nu_w = 0$. Q.E.D.

## EC.7 Allowing the Provider to Shirk

In the basic setup of our paper, we focus on cases where the provider either works or not. In this section, we consider the case where the firm lets the provider to shirk, and analyze the effect of improving the acquisition capacity on the firm's profit. We first derive the expected profit of the firm when the provider shirks in one period but works in the other.

Lemma EC.3. *The firm's optimal expected profit when the provider shirks in the first and works in the second period is given by*

$$V_{sw}(d_0) = \pi_s(d_0)2R_\nu(\delta^{\mathtt{M}},\alpha_h) + \bar{\pi}_s(d_0)\left[\max_{\delta\in[0,\nu]} R_\nu(\delta,\alpha_\ell) + \pi_w(d_0+\delta)\tau(\alpha_h-\alpha_\ell) - \frac{\kappa}{\rho(d_0+\delta)} + \tau\alpha_\ell\right].$$

*The firm's optimal expected profit when the provider works in the first and shirks in the second period is given by*

$$V_{ws}(d_0) = \pi_w(d_0)2R_\nu(\delta^{\mathtt{M}},\alpha_h) + \bar{\pi}_w(d_0)\left[\max_{\delta\in[0,\nu]} R_\nu(\delta,\alpha_\ell) + \pi_s(d_0+\delta)\tau(\alpha_h-\alpha_\ell) + \tau\alpha_\ell\right] - \frac{\kappa}{\rho(d_0)}.$$

*Proof of Lemma EC.3.* If the provider works only in the second period, the probability of observing $\alpha_h$ in the first period is $\pi_s(d_0)$. In this case, the firm collects the myopic optimal revenue for two periods to get $R_\nu(\delta^{\mathtt{M}},\alpha_h)$ (see Lemma 1). With the remaining probability $\bar{\pi}_s(d_0)$, the firm observes $\alpha_\ell$ in the first period. Then, the firm needs to collect data by considering the current sales revenue, the AI Flywheel effect (with probability $\pi_w(d_0+\delta)$ for a given $\delta$), and the incentive issues. The moral hazard problem is reduced to a single period problem with high accuracy probabilities $\pi_e(d_0+\delta)$ for $e \in \{s,w\}$ for a given $\delta$. Thus, the firm's expected payment to the provider in the second period is given by $\kappa/\rho(d_0+\delta)$ for a given $\delta$ (see Proposition 4.2 in Laffont and Martimort 2009, p. 157). Combining all these terms, the firm's profit is given by

$$V_{sw}(d_0) = \pi_s(d_0)2R_\nu(\delta^{\mathtt{M}},\alpha_h) + \bar{\pi}_s(d_0)\left[\max_{\delta\in[0,\nu]} R_\nu(\delta,\alpha_\ell) + \pi_w(d_0+\delta)\tau(\alpha_h-\alpha_\ell) - \frac{\kappa}{\rho(d_0+\delta)} + \tau\alpha_\ell\right].$$

If the provider shirks in the second period, the moral hazard problem is with probabilities $\pi_e(d_0)$ for $e \in \{s,w\}$, and the firm's data collection decision needs to consider only the current sales revenue

and the AI Flywheel effect (with probability $\pi_s(d_0+\delta)$ for a given $\delta$). Hence, we obtain the optimal profit as follows.

$$V_{ws}(d_0) = \pi_w(d_0)2R_\nu(\delta^{\mathtt{M}}, \alpha_h) + \bar{\pi}_w(d_0)\left[\max_{\delta\in[0,\nu]} R_\nu(\delta,\alpha_\ell) + \pi_s(d_0+\delta)\tau(\alpha_h-\alpha_\ell) + \tau\alpha_\ell\right] - \frac{\kappa}{\rho(d_0)}\,.$$

Note also that the optimization problems in both expressions have unique solutions because the objective functions are concave and the set $[0,\nu]$ is closed and bounded. Q.E.D.

In the following proposition, we first characterize how $V_{ws}(d_0)$ changes in $\nu$.

Proposition EC.14. *The expected profit $V_{ws}(d_0)$ is strictly increasing in $\nu$.*

*Proof of Proposition EC.14.* We directly derive the expressions for these derivatives. Let $q^{ws}$ and $\delta^{ws}$ denote the optimal quantities and data collection respectively for this setting.

$$\begin{aligned}\frac{\partial V_{ws}(d_0)}{\partial \nu} =& \bar{\pi}_w(d_0)\left\{\frac{\partial q^{ws}}{\partial \nu}\left[\alpha_\ell(1-2q^{ws}) + \pi_s'(d_0+q^{ws}\nu)\tau(\alpha_h-\alpha_\ell)\nu\right] + \pi_s'(d_0+q^{ws}\nu)\tau(\alpha_h-\alpha_\ell)q^{ws}\right\}\\ =& \bar{\pi}_w(d_0)\frac{\alpha_\ell}{\nu}(2q^{ws}-1)q^{ws} = \bar{\pi}_w(d_0)\frac{\alpha_\ell}{\nu}\left(\frac{2\delta^{ws}}{\nu}-1\right)\frac{\delta^{ws}}{\nu} > 0\,.\end{aligned}$$

Here, the second equality follows from the algebraic simplifications based on the first-order condition which $q^{ws}$ satisfies. The third inequality follows from $\delta = \nu q$. Finally, the last inequality follows from the fact that $\delta^{ws} > \nu/2$ which is the myopic data collection. This result can be proved by following verbatim the same steps in Proposition 2 using $\pi_s(d_0+\delta)$ instead of $\pi_w(d_0+\delta)$. Q.E.D.

Proposition EC.15. *We have:*

1. *If $\rho(\cdot)$ is strictly increasing, the expected profit $V_{sw}(d_0)$ is strictly increasing in $\nu$.*
2. *If $\rho(\cdot)$ is strictly decreasing, a unique threshold $\nu_{sw}$ exists such that the expected profit $V_{sw}(d_0)$ is strictly increasing in $\nu$ if $\nu \le \nu_{sw}$, and is strictly decreasing otherwise. Further, threshold $d_{sw} > 0$ exists such that the profit is strictly decreasing (i.e., $\nu_{sw} = 0$) if $d_0 \ge d_{sw}$.*

*Proof of Proposition EC.15.* Note that the optimization problem inside the expected profit $V_{sw}(d_0)$ is the same as (EC.8). Thus, we can use Proposition EC.8 to show that the same steps in the proof of Theorem 4 can be followed in this setup, too. Q.E.D.

We conclude this section by noting that Theorems 1-3 can be repeated verbatim for the setup where the provider shirks in the first period and works in the second period using (EC.8) as discussed in Section EC.4 of this e-companion.

## Recent ESMT Working Papers

| | ESMT No. |
|---|---|
| **Beyond retail stores: Managing product proliferation along the supply chain**<br>Işık Biçer, SchulichSchool of Business, YorkUniversity<br>Florian Lücker, Cass Business School, City, University of London<br>Tamer Boyaci, ESMT Berlin | 19-02 (R2) |
| **Informing the public about a pandemic**<br>Francis de Véricourt, ESMT Berlin<br>Huseyin Gurkan, ESMT Berlin<br>Shouqiang Wang, Naveen Jindal School of Management, The University of Texas at Dallas | 20-03 (R2) |
| **Human and machine: The impact of machine input on decision-making under cognitive limitations**<br>Tamer Boyaci, ESMT Berlin<br>Caner Canyakmaz, ESMT Berlin<br>Francis de Véricourt, ESMT Berlin | 20-02 |
| **Queueing systems with rationally inattentive customers**<br>Caner Canyakmaz, ESMT Berlin<br>Tamer Boyaci, ESMT Berlin | 18-04 (R1) |
| **The impact of EU cartel policy reforms on the timing of settlements in private follow-on damages disputes: An empirical assessment of cases from 2001 to 2015**<br>Hans W. Friederiszick, ESMT Berlin and E.CA Economics<br>Linda Gratz, E.CA Economics<br>Michael Rauber, E.CA Economics | 19-03 (R1) |