WORKING PAPERS

N° 1621

February 2025

"Digital Ecosystems and Data Regulation"

Andrew Rhodes, Jidong Zhou and Junjie Zhou

Toulouse School of Economics

Digital Ecosystems and Data Regulation*

Andrew Rhodes† Jidong Zhou‡ Junjie Zhou§

February 2025

Abstract

This paper provides a framework in which a multiproduct ecosystem competes with many single-product firms in both price and innovation. The ecosystem is able to use data collected on one product to improve the quality of its other products. We study the impact of data regulation which either restricts the ecosystem's cross-product data usage, or which requires it to share data with small firms. Each policy induces small firms to innovate more and set higher prices; it also dampens data spillovers within the ecosystem, reduces the ecosystem's incentive to collect data and innovate, and potentially increases its prices. As a result, data regulation has an ambiguous impact on consumers, and is more likely to benefit consumers when small firms are relatively more efficient in innovation. A data cooperative among small firms, which helps them to share data with each other, does not necessarily benefit small firms and can even harm consumers.

Keywords: digital ecosystems, innovation, data regulation, data cooperative

JEL classification: D43, L13, L51

---

*We are grateful to Nageeb Ali, Mark Armstrong, Yeon-Koo Che, Jay Pil Choi, Matt Elliott, Sanjeev Goyal, Meg Meyer, Guofu Tan, Tat-How Teh, Julian Wright, and various seminar and conference audiences for helpful comments. We thank Chengqing Li for his excellent research assistance. Rhodes acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program (grant ANR-17-EURE-0010) and funding from the European Union (ERC, DMPDE, grant 101088307).

†Toulouse School of Economics, Toulouse University Capitole. andrew.rhodes@tse-fr.eu

‡School of Management, Yale University. jidong.zhou@yale.edu

§School of Economics and Management, Tsinghua University. zhoujj03001@gmail.com

1 Introduction

There are widespread concerns that big tech companies like Google, Apple, Meta, Amazon, and Microsoft have become too big and too powerful.[1] These companies operate as "digital ecosystems" that offer a very large range of products and services. They generate vast amounts of data, and can use data gleaned in one product market to improve their offering in other markets.[2] Moreover, these ecosystems typically compete not just with each other, but also with smaller firms that specialize in particular products, and which lack access to the same volume or scope of data.

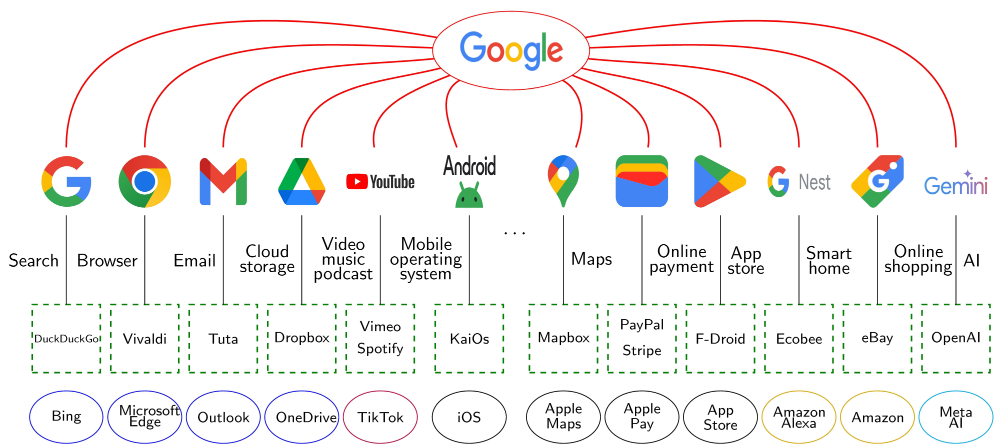


Figure 1: Google's ecosystem and some of its competitors
(specialized competitors in boxes; competing products from other ecosystems in ovals)

To illustrate, Figure 1 depicts some of the diverse businesses operated by Google, as well as some of its competitors. There are many ways in which Google can leverage data across these different businesses. For example, it can use spending data from Google Pay, location data from Google Maps, as well as data on consumer trends from Google Shopping, to enhance the relevance of its search and advertising results, in a

[1]According to a 2023 survey, 60% of U.S. respondents felt that big tech companies "have too much power in the market, which puts competitors at a disadvantage and hurts both smaller businesses and consumers." (See https://shorturl.at/PQVs2.) Similar concerns have been raised about big tech companies like Tencent, Alibaba, and ByteDance in China. (See, e.g., https://shorturl.at/Z7WRu.)

[2]For example, Google's privacy policy says "We use automated systems that analyze your content to provide you with things like customized search results, personalized ads, or other features tailored to how you use our services. ... We may use the information we collect across our services and across your devices for the purposes described above." (See https://shorturl.at/3uIb8.)

way that a specialized competitor like DuckDuckGo cannot. Similarly, it can use data from sources like Google search and YouTube to train its AI chatbot, Gemini, enabling it to catch up with—and potentially surpass—its specialized competitor, OpenAI.[3] Meanwhile, ecosystems like Tencent and Alibaba that provide fintech lending services, can use data from social apps (e.g., WeChat) or e-commerce sites (e.g., Taobao) to improve the accuracy of default-risk predictions and make faster lending decisions.

The ability of ecosystems to leverage data across their different business units has sparked fears that specialized firms may be placed at a competitive disadvantage, dampening their incentives to innovate and expand, and ultimately harming final consumers.[4] At the same time, some recent legislation may affect ecosystems' data advantage. For instance, privacy policies like the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) could weaken ecosystems' data advantage if they induce consumers to share less data with big tech companies. Meanwhile, the Digital Markets Act (DMA) prohibits large ecosystems from combining data across markets without explicit user consent—reducing the extent of cross-market data usage within ecosystems.[5] The DMA also facilitates data sharing with smaller competitors by enabling data interoperability, further reducing ecosystems' relative data advantage.[6] More broadly, initiatives like Gaia-X, which aim to enable data sharing among individual firms in Europe, may also empower small firms to compete more effectively with large digital ecosystems.

[3]See, for example, https://shorturl.at/lDuSH and https://shorturl.at/psQLq on the importance of data for AI and how Google's data advantage helps Gemini compete with ChatGPT.

[4]For instance, according to ACCC (2023), "competitors who do not have access to the same volume or scope of consumer data may find themselves at a competitive disadvantage relative to the digital platform ecosystem." Our paper focuses on ecosystems' data advantage, but we acknowledge that other factors—such as financial resources, brand recognition, and superior infrastructure—also give them a competitive advantage over smaller firms.

[5]Firm-specific regulation can also affect cross-market data usage by ecosystems. For example, due to regulatory pressure, Ant Group—formerly Alibaba's financial services division—became a more independent entity and terminated its data-sharing agreement with Alibaba in 2022. (See https://shorturl.at/Wd8im.) Also, in 2024 the Indian Competition Commission banned WhatsApp from sharing user data with other entities owned by its parent company Meta. (The ban was suspended in January 2025. See https://shorturl.at/3EIBd.)

[6]Article 5(2) of the DMA concerns combining data from different businesses, while Article 6(9) concerns data portability.

However, it is unclear how these data policies will affect competition between ecosystems and specialized firms. For example, will specialized firms innovate more and benefit from these policies? How will market prices and the ecosystem's incentive to innovate change? Will these policies eventually help consumers? In this paper we develop a framework to address these questions.

In Section 2 we lay out a model in which a digital ecosystem operates in a large number of product markets, and in each market competes with a different single-product firm. Firms compete in both price and innovation/quality investment; this enables us to study how regulation affects innovation incentives—a central concern for policymakers.[7] A key feature of our model is that the ecosystem can leverage data collected in one market to enhance the quality of its product in another market. This "data spillover" effect generates demand complementarity across products. We analyze the model in Section 3. Given the data-driven demand complementarities, we face similar technical issues concerning equilibrium existence and multiplicity as does the literature on competition with network effects.[8] Our problem is even more challenging, though, due to the additional innovation choice, rich product heterogeneity, and the asymmetric competition between a multiproduct firm and single-product firms. Despite these challenges, we establish conditions for the existence and uniqueness of an interior equilibrium in which both firms in each market make positive sales. As one might expect, the ecosystem has an incentive to set relatively low prices in order to accumulate more data. However, surprisingly, the ecosystem does not necessarily set lower prices on those products which generate more data. Intuitively, if the ecosystem is efficient enough at innovating, it prefers to expand sales of those products by investing in higher product quality, which then allows it to also charge higher prices for them.

In Section 4 we use our framework to examine the impact of data regulation. We consider two policies designed to level the playing field between the ecosystem and small firms: firstly, restricting cross-market data usage, which limits the ecosystem's ability to use data collected from one business to gain a competitive advantage in another, and

[7]For instance, the DMA aims "to ensure a contestable and fair digital sector ... with a view to promoting innovation ... as well as a high quality and choice for end users..."

[8]See, e.g., Katz and Shapiro (1985) and Chapter 7.8 of Anderson, de Palma, and Thisse (1992). Our setup is closer to the latter which studies price competition with both horizontal product differentiation and network effects. However, we are not aware of any theory papers in this literature where asymmetric firms compete in both price and quality investment.

secondly, requiring the ecosystem to share data with smaller rivals.[9,10] Each policy helps small firms by inducing them to innovate more and sell more. However, these policies do not necessarily benefit consumers, and can sometimes even benefit the ecosystem. The perverse effect on consumers can occur for three reasons: first, while small firms innovate more under regulation, they also set higher prices; second, the ecosystem innovates less under regulation, but its prices do not necessarily decrease to compensate its quality decline, as the regulation dampens its incentive to acquire data by offering low prices; third, the regulation hampers the ecosystem's ability to use data to improve its product quality. Data regulation is more likely to benefit consumers when small firms are more efficient in innovation, since in this case their investment responds more strongly to the data policies. (If firms compete only in price, the regulation induces no innovation response and so is unambiguously bad for consumers.) Finally, we show that at the margin a policy that encourages data sharing Pareto dominates a policy that restricts the ecosystem's cross-product data usage.

In Section 5 we consider the possibility that single-product firms share data with each other via a "data cooperative". Small firms face a trade-off when establishing such a cooperative. On the one hand, they can use each other's data to improve their product quality and hence reduce the ecosystem's data advantage. On the other hand, they induce the ecosystem to price more aggressively, because as the ecosystem sells to more consumers it not only generates more data for itself, but now it also deprives the small firms of data. When products are symmetric, small firms benefit from forming a data cooperative (i.e., the first effect dominates the second) if and only if they already hold a larger market share than the ecosystem. Consumers always benefit from a data cooperative when products are sufficiently symmetric to each other, but otherwise can suffer due to the reduced innovation by the ecosystem.

Finally, we note that the demand complementarity across products caused by data spillovers in our model can also be interpreted in more traditional ways. For instance, a multiproduct firm might benefit from cross-product network effects, where having more

[9]In practice, data sharing requires either compensation (if firms own the data) or consumer consent (if the data belongs to consumers). For example, the latter approach is adopted in the recent open banking policy in Europe, which mandates that traditional banks share their data with new entrants, such as fintech lenders, if consumers consent. (See, e.g., He, Huang, and Zhou, 2023.)

[10]In our framework policies that restrict how much data firms can collect, such as GDPR, have qualitatively the same impact as these two data regulation policies.

consumers for one product enhances the utility of its other products. Alternatively, it might benefit from learning-by-doing across products, where selling or producing in one market helps it learn to perform better in other markets. However, under these alternative interpretations, our analysis of data regulation and data cooperatives—the key policy implications of the model—lack practically relevant counterparts.[11]

1.1 Related Literature

Data-driven product improvement/innovation. Several recent empirical papers have documented evidence that customer data can help improve product quality. For example, access to more data improves answers to queries on search engines (Yoganarasimhan, 2020; Schaefer and Sapi, 2023), and improves recommendations for online news (Peukert, Sen, and Claussen, 2024). There is also evidence on data-driven innovation. For instance, access to government data increases commercial software development in China's facial recognition AI industry (Beraja, Yang, and Yuchtman, 2023), and data sharing among app developers in China boosts innovation (Zhou, 2025).

There are also theoretical papers that explore the implications of data-driven product improvement.[12] For instance, Prüfer and Schottmüller (2021) and Hagiu and Wright (2023) study dynamic duopoly models where customer data from previous transactions can be used to improve product quality. In Prüfer and Schottmüller (2021), two firms compete in quality choice at each period, and their costs of producing quality decrease in the number of customers served in the previous period. They show that there is a strong tendency for the market to tip, but that regulation which forces firms to share data with each other can prevent this tipping from occurring (and induce both firms to choose higher qualities). In Hagiu and Wright (2023), two firms compete in price, instead of quality, à la Bertrand, and in each period the value of a firm's product is increasing in the number of its past users. They also find that the market tips in favor

[11] Note also that our paper does not consider the possibility that the ecosystem's cross-market data usage directly harms consumers (e.g., via price discrimination). Instead, we focus on how the ecosystem's data advantage might indirectly harm consumers by weakening smaller specialized competitors.

[12] A related idea has also recently been explored in the macro literature. For example, Farboodi, Mihet, Philippon, and Veldkamp (2019) assume that data improves firms' forecast accuracy and so their production decisions; Jones and Tonetti (2020) assume that data improves the quality of firms' ideas and so their production efficiency.

of one firm, and firms may price below cost to generate data in order to induce this market tipping. A policy that forces an incumbent to share data with an entrant can then harm consumers, since it dampens firms' incentives to compete for data.

Compared to these papers, we focus on cross-market data usage by a multiproduct firm competing with single-product rivals. While we do not address dynamic data accumulation, we allow firms to compete in both price and quality. Similar to Hagiu and Wright (2023), data sharing in our model can soften price competition, but it also affects innovation. We also examine additional policies, such as data cooperatives.

Digital ecosystems. There is an emerging economics literature on digital ecosystems.[13] Condorelli and Padilla (2024) propose an entry deterrence theory of digital ecosystems. In their model, initially firm A operates in a primary market which uses data, and firm B operates in a secondary market which generates data. If firm A enters the secondary market (i.e., becomes an ecosystem), it can use the data generated there to deter firm B from entering the primary market. The paper discusses several policies which can mitigate entry deterrence, such as a ban on privacy-policy tying (to prevent cross-market data usage), and data portability between competing firms.

Heidhues, Köster, and Kőszegi (2024) offer a conglomerate-merger theory for the formation of digital ecosystems (or multiproduct firms in general). They introduce a default effect in consumer choice: after purchasing a product from a multiproduct firm in the primary market, a consumer regards its product in the secondary market as the default option and tends to choose it over competing products. Then a single-product firm that sells more in the primary market has a stronger incentive to leverage this default effect by acquiring firms in the secondary market (and so become an ecosystem).

Unlike these two papers, we do not consider the formation of an ecosystem. We focus instead on how an existing ecosystem influences innovation and expansion by its small rivals, and how this is affected by data policies. We do this using a fully-fledged model of competition (in both price and innovation), which is absent in these papers.

Kraemer and Shekhar (2024) study closely related questions using a different setup,

[13] Platforms like e-commerce websites or app stores are also sometimes considered digital ecosystems. For instance, Bisceglia and Tirole (2024) take this perspective and examine issues such as excessive platform fees and self-preferencing by gatekeeper platforms that also sell their own products or services. Anderson and Bedre Defolie (2025) examine the interplay between platform commissions, device fees, and entry decisions of app providers, including when some apps belong to the platform.

in which an ecosystem monopolizes a primary market, and uses data from it to improve its product in a secondary market, where it competes à la Cournot with another firm. Firms choose innovation in each market and are equally efficient at it. Like us, the authors study policies that restrict cross-market data usage or force the ecosystem to share data, but find different welfare impacts.[14] This is mainly because we study a richer model with price competition where an ecosystem competes in many markets, each of which can generate and use data, and firms can differ in their innovation efficiency.

Jeon, Lefouili, Li, and Simcoe (2024) study a multiproduct upstream supplier (interpreted as an ecosystem) that sells a number of inputs. The price of each downstream product equals the total input price, while demands for these products are linear and exhibit cross-product externalities. The authors show how an input's optimal monopoly price depends on its centrality in the externality network. Our competition model is quite different, as is our focus on the impact of data regulation.

Multiproduct vs single-product firms. Asymmetric competition between a multiproduct firm and several single-product firms is rarely studied theoretically in the literature. One exception is some recent work on conglomerate mergers, such as Rhodes and Zhou (2019) and Chen and Rey (2023).[15] In both papers, two firms initially operating in separate markets can merge into a multiproduct firm and then compete with the remaining single-product firms in each market. Such a merger can be profitable either due to one-stop shopping convenience when consumers face search frictions (in Rhodes and Zhou, 2019) or because of consumption synergies (in Chen and Rey, 2023). These papers, however, focus only on price competition, and do not have data-driven cross-product externalities, and hence address very different research questions to us.

Multi-sided platforms. Due to data spillovers across products, the ecosystem in our model can also be regarded as a multi-sided platform with cross-market network effects. The existing (theoretical) works in that literature, however, usually focus on *symmetric* competitive platforms for tractability. See, e.g., Armstrong (2006), Rochet

[14] In their model the first policy harms consumers while the second one benefits them, and both policies hurt the ecosystem. However, in our model data regulation can either benefit or harm consumers and the ecosystem, depending on firms' relative innovation efficiency and product heterogeneity.

[15] A related but different asymmetric market structure is considered in the literature on the leverage theory of bundling (e.g., Whinston, 1990, and Nalebuff, 2004). In those models a multiproduct firm that faces a single-product entrant in one market has an incentive to use a bundling strategy so as to leverage its monopoly power and hence foreclose the potentially competitive market.

and Tirole (2006), Tan and Zhou (2021), and the survey paper by Jullien, Pavan, and Rysman (2021). One exception is Peitz and Sato (2023), who study price competition among two-sided platforms that can differ in their within- and cross-group network effects, costs, and exogenously given qualities; they show that the model is tractable when taste shocks are logistic, and network effects have a logarithmic specification.[16] We also allow for rich heterogeneity across firms and products, but study a different type of asymmetric competition between a large ecosystem and many single-product firms. Our focus is also more on the welfare effects of various data policies. Finally, we note that innovation/quality choice plays an important role in our analysis, but is not usually studied in the literature on multi-sided platforms (including the above papers).

2 The Model

A digital ecosystem competes with single-product firms in a continuum of product markets. (As we will explain later, this continuum assumption simplifies the subsequent analysis.) To focus on the role of data, we assume that these products are intrinsically independent of each other. Without loss of generality, denote by $\mathcal{I} = [0, 1]$ the set of all products, and endow it with the standard Lebesgue measure di. Each product $i \in \mathcal{I}$ is supplied by the ecosystem and a different single-product firm; they compete for consumers in that market by simultaneously choosing both price and quality/innovation. Let $p_{e,i}$ and $v_{e,i}$ denote the ecosystem's price and quality investment on product i, and let $p_{s,i}$ and $v_{s,i}$ denote the single-product firm i's price and quality investment.[17] We assume that the fixed costs of investment for product i are given by respectively

$$C_{e,i}(v_{e,i}) = \frac{v_{e,i}^2}{2\eta_{e,i}} \quad \text{and} \quad C_{s,i}(v_{s,i}) = \frac{v_{s,i}^2}{2\eta_{s,i}}.$$

We allow for asymmetries in firms' innovation costs. For example, if small firms are more efficient at innovating than the ecosystem in market i due to their specialization,

[16] See also Belleflamme, Peitz, and Toulemonde (2022) for a model of asymmetric platforms with linear Hotelling differentiation and linear network effects. Empirical research on competitive multi-sided platforms (e.g., Rysman, 2004), however, allows for firm asymmetry, since analytical tractability is not a concern. Sometimes they also consider endogenous quality choices (see, e.g., Fan, 2013).

[17] In some markets firms do not charge consumers for using their services, and instead make profit from advertising. In that case we can interpret a firm's price as a proxy for the amount of ads it displays; displaying more ads causes more disutility to consumers but yields more revenue for the firm.

we would have $\eta_{s,i} > \eta_{e,i}$; if instead the ecosystem is more efficient, we would have $\eta_{e,i} > \eta_{s,i}$. However, we assume that both the ecosystem and single-product firms have a constant marginal production cost, which is independent of product quality, and which for convenience we normalize to zero. (We note that for many digital services, the marginal cost is negligible, and most costs are from developing or improving services.)

The ecosystem generates data from its customers. We assume that each market has a unit mass of consumers and is fully covered (i.e., each consumer buys from one of the two firms). (It does not matter for our analysis if the same consumers are present in each market, or if different markets have different consumers.[18]) The total amount of data generated by the ecosystem is

$$Q = \int_{\mathcal{I}} \alpha_i z_i di, \tag{1}$$

where $\alpha_i \geq 0$ measures how good product i is at generating data, and $z_i \in [0, 1]$ denotes the fraction of consumers who buy product i from the ecosystem. We assume that all the integrals in this paper are well defined. (From now on we omit $\mathcal{I}$ in the integral when no confusion arises.)

Data is used to improve products: if a consumer buys product i from the ecosystem, her surplus is

$$v_{e,i} - p_{e,i} + \beta_i Q + \epsilon_{e,i};$$

whereas if she buys from single-product firm i, her surplus is

$$v_{s,i} - p_{s,i} + \epsilon_{s,i}.$$

The parameter $\beta_i \geq 0$ measures how much the ecosystem can use data to improve a consumer's utility for its product i, beyond the quality improvement from its innovation investment.[19] Note that in this setup with a continuum of products, the amount of data each single-product firm has is negligible compared to what the ecosystem has, and so

[18] We can also allow markets to differ in their size. If market i has a measure m_i of consumers, our later analysis applies provided we replace the product space measure di by $m_i di$, and replace $(\eta_{e,i}, \eta_{s,i})$ by $(m_i\eta_{e,i}, m_i\eta_{s,i})$. (The latter is because firms invest more in a larger market, other things equal.)

[19] For simplicity, we assume linear data spillovers and quadratic cost functions. Considering more general spillovers and cost functions does not change the basic logic of our analysis, but establishing equilibrium existence and uniqueness is more challenging, and comparative statics are less tractable.

we assume that it has no impact on their product quality. We allow products to differ arbitrarily in how much data they generate α_i, and how well they use data β_i.

The taste shocks $\epsilon_{e,i}$ and $\epsilon_{s,i}$ capture idiosyncratic preferences for product i. Given the assumption of full market coverage, only the difference in taste shocks $\epsilon_{s,i} - \epsilon_{e,i}$ matters for consumer choices. We assume that for a given product market i, the difference in taste shocks $\epsilon_{s,i} - \epsilon_{e,i}$ is i.i.d. across consumers according to a distribution F_i. Therefore, in each market consumer choice is represented by a Hotelling model with a general preference distribution.[20] We further assume that F_i has a differentiable pdf f_i which is log-concave and symmetric around 0 on support $[-l_i, l_i]$ (where l_i can be infinity). That is, for each product there is symmetric product differentiation between the ecosystem and the single-product firm.[21]

Finally, the timing is as follows. Firms simultaneously choose prices and quality investments in each market, consumers in each market observe those choices and form a rational expectation about Q, and then decide which product to buy.

Modeling discussions. Before proceeding with the analysis, we discuss some of our modeling choices:

(i) *A continuum of products.* Considering a continuum of products/markets is an approximation of the fact that ecosystems usually operate in a large number of markets. It also makes our model significantly more tractable, because it implies that a single-product firm does not affect how much data the ecosystem generates. If instead we had a finite number of markets, each single-product firm would affect Q and hence the profits of other single-product firms; this would create non-trivial strategic interactions among these small firms. In the Online Appendix, we illustrate this additional complexity in the two-product case.

(ii) *Two channels of quality improvement.* In our model, product quality can be improved via two channels: innovation investment and data-driven enhancements. The

[20] Note that if the same consumer is present in multiple markets, it does not matter for our analysis whether her $\epsilon_{s,i} - \epsilon_{e,i}$ are independent across markets or not provided F_i is the marginal distribution.

[21] We obtain qualitatively the same results if we allow for asymmetric product differentiation, such that f_i is symmetric around some $\hat{x}_i \neq 0$. If the same set of consumers is present in multiple markets, this more general setup can also help capture one-stop-shopping convenience offered by the ecosystem. Specifically, suppose that if a consumer buys from a single-product firm, she needs to pay some extra cost; this is the same as shifting the product-differentiation distribution towards the left in each market.

former reflects efforts such as developing new service features or recommendation algorithms, while the latter captures how data can be used to improve existing services or algorithms. Note that we can shut down the innovation channel by setting all innovation efficiency parameters (i.e., the η's) to zero. However, including the innovation channel allows us to examine how the ecosystem's data advantage and related data policies influence firms' innovation incentives—–a key concern for policymakers. Moreover, having this innovation channel can qualitatively affect the impact of data regulation.

(iii) *Data and innovation cost.* Access to more data could make innovation less costly, rather than improving product quality directly as in our model. As we demonstrate in the Online Appendix, this alternative model can be solved using a similar approach as the one we employ below, and our main insights are robust.

(iv) *More general cross-product data spillovers.* We could consider a more general model in which the ecosystem's product i offers surplus $v_{e,i} - p_{e,i} + \int \beta_{ji} F_j(\Delta_j) dj + \epsilon_{e,i}$, where β_{ji} denotes how much the data from a unit of sales of product j helps improve product i. (Our current setup is the case with $\beta_{ji} = \alpha_j \beta_i$.) This general setup allows for, e.g., the possibility that $\beta_{ji} > \beta_{ji'}$ but $\beta_{ki} < \beta_{ki'}$, i.e., product j's data is more useful for product i than for product i' but the opposite is true for product k's data. As shown in the Online Appendix, our analysis can be generalized to this more general case. However, one advantage of our current simpler approach is that we can talk about data generation and data usage separately for each product.

3 Equilibrium Analysis

We now solve the model, starting with the consumer problem, before moving to firms' choices of price and investment. We then provide conditions for existence and uniqueness of equilibrium.

3.1 Consumer Problem

Let the basic surplus difference between the two products in market i be

$$\Delta_i \equiv v_{e,i} - p_{e,i} + \beta_i Q - (v_{s,i} - p_{s,i}). \tag{2}$$

Since by assumption each market is fully covered, a consumer in market i buys the ecosystem's product i if

$$\Delta_i \geq \epsilon_{s,i} - \epsilon_{e,i}$$

and otherwise buys from the single-product firm i. Demand for the ecosystem's product i is therefore $z_i \equiv F_i(\Delta_i)$, and demand for the single-product firm i is $1 - F_i(\Delta_i)$. For given prices and qualities, the equilibrium of the consumer choice game with rational expectations is then characterized by Q which solves

$$Q = \int \alpha_i F_i(\Delta_i) di, \tag{3}$$

where we have used the earlier equation (1). Using the observation that f_i is symmetric and single-peaked, and hence $f_i(0) \geq f_i(\Delta_i)$, we obtain the following result:

Lemma 1. *Define $\mathbb{E}[\alpha] \equiv \int \alpha_i di$. Then, for any given prices and qualities, the consumer choice game has a unique (stable) equilibrium $Q \in [0, \mathbb{E}[\alpha]]$ if*

$$\mathbb{E}[\alpha\beta] \equiv \int \alpha_i \beta_i di < \frac{1}{\max_i f_i(0)}. \tag{4}$$

This is simply because (4) ensures that the right-hand side of (3) increases in Q less quickly than does the left-hand side. We note that the condition in (4) is implied by our later Assumption 1. In some special cases (e.g., when each F_i is a uniform distribution), one can explicitly solve Q as a function of all prices and qualities; in general, however, (3) does not have an analytical solution.

It is easy to see that, due to data spillovers, there is demand complementarity across the ecosystem's products. In particular, if the ecosystem lowers price or invests more in a subset of markets where $\alpha_i > 0$, then under condition (4) this leads to a higher Q, which then leads to higher ecosystem demand in *all* markets for which $\beta_i > 0$.

3.2 Firm Problem

We now turn to the firms' optimization problems. To avoid trivial cases, from now on we assume that there is a strictly positive measure of products that generate strictly positive data (i.e., have $\alpha_i > 0$), and a strictly positive measure of products where the ecosystem uses data to improve their quality (i.e., have $\beta_i > 0$).

The advantage of considering a continuum of products is that a single-product firm's choice of price and quality has no effect on the ecosystem's total data Q, and hence also no effect on other single-product firms. This implies that there is no strategic interaction among single-product firms, which significantly simplifies the analysis.

Single-product firm i's problem is to

$$\max_{p_{s,i},v_{s,i}} \; p_{s,i}[1 - F_i(\Delta_i)] - C_{s,i}(v_{s,i}) \tag{5}$$

by taking Q in Δ_i as fixed. On the other hand, the ecosystem's problem is to

$$\max_{\{p_{e,i},v_{e,i}\}_{i\in\mathcal{I}}} \int [p_{e,i}F_i(\Delta_i) - C_{e,i}(v_{e,i})]di \tag{6}$$

where Q in Δ_i solves (3).

In the following, we focus on an *interior* equilibrium where both firms in each market have some demand. We first report the first-order conditions for an interior equilibrium, and then provide conditions for the existence and uniqueness of such an equilibrium. We can solve each firm's optimization problem using two different approaches. One is to work with the system of price and quality, which is more convenient in characterizing the equilibrium; the other is to work with the system of demand quantity and quality, which turns out to be more convenient in investigating existence of equilibrium.

Start with single-product firm i's problem in (5). The first-order conditions with respect to $p_{s,i}$ and $v_{s,i}$ yield respectively

$$\underbrace{1 - F_i\left(\Delta_i\right)}_{\text{MC of price cut}} = \underbrace{p_{s,i}f_i(\Delta_i)}_{\text{MB of price cut}} \quad \text{and} \quad \underbrace{v_{s,i}/\eta_{s,i}}_{\text{MC of investment}} = \underbrace{p_{s,i}f_i(\Delta_i)}_{\text{MB of investment}}.$$

Intuitively, in order to sell to more consumers, single-product firm i can either reduce its price or increase its investment. In both cases the marginal benefit is the same—the resulting increase in sales $f_i(\Delta_i)$ multiplied by the price $p_{s,i}$. The marginal "cost" of cutting price is the demand $1 - F_i\left(\Delta_i\right)$ on which the price cut accrues, while the marginal cost of raising investment is $v_{s,i}/\eta_{s,i}$. At the optimum these costs must be equal. Jointly solving the two first-order conditions gives:

$$p_{s,i} = \frac{1 - F_i(\Delta_i)}{f_i(\Delta_i)} \quad \text{and} \quad v_{s,i} = \eta_{s,i}[1 - F_i(\Delta_i)]. \tag{7}$$

Price takes the standard form, and is equal to demand divided by demand sensitivity. Meanwhile investment is proportional to demand because, as explained above, at the optimum the marginal "costs" of cutting price or raising investment are the same.

Now consider the ecosystem's problem in (6). This is more complicated because it is an infinite-dimensional optimization problem, and because the ecosystem can affect the amount of data Q that it generates. Since Q cannot usually be solved analytically from equation (3), we use the Lagrangian method and treat Q as a choice variable:

$$\mathcal{L} = \int \left[p_{e,i} F_i(\Delta_i) - \frac{v_{e,i}^2}{2\eta_{e,i}} \right] di + \lambda \left[\int \alpha_i F_i(\Delta_i) di - Q \right],$$

where the Lagrange multiplier λ captures the marginal value to the ecosystem of generating extra data. The first-order conditions with respect to $p_{e,i}$ and $v_{e,i}$ yield respectively

$$\underbrace{F_i(\Delta_i)}_{\text{MC of price cut}} = \underbrace{[p_{e,i} + \lambda\alpha_i] f_i(\Delta_i)}_{\text{MB of price cut}} \quad \text{and} \quad \underbrace{v_{e,i}/\eta_{e,i}}_{\text{MC of investment}} = \underbrace{[p_{e,i} + \lambda\alpha_i] f_i(\Delta_i)}_{\text{MB of investment}}.$$

The explanation is the same as for the single-product firms' first-order conditions, except that the marginal benefit terms are different. In particular, due to cross-product data usage, each additional unit of sales on product i now not only generates a direct revenue $p_{e,i}$, but also raises the amount of data by α_i, which the ecosystem values at λ. Jointly solving the first-order conditions gives:

$$p_{e,i} = \frac{F_i(\Delta_i)}{f_i(\Delta_i)} - \lambda\alpha_i \quad \text{and} \quad v_{e,i} = \eta_{e,i} F_i(\Delta_i). \tag{8}$$

The ecosystem's price is qualitatively different to that of a single-product firm: it has an extra term $-\lambda\alpha_i$, reflecting the ecosystem's additional incentive to cut prices so as to accumulate more data. This incentive is stronger on products that generate more data (i.e., have larger α_i) and when data is more valuable (i.e., λ is larger). However, the ecosystem's investment is still proportional to its demand, and so is qualitatively the same as that of a single-product firm. Intuitively, even though the ecosystem has more incentive to expand demand due to cross-product spillovers, it faces the same trade-off at the margin between doing this by cutting price or raising investment.

In order to pin down the ecosystem's price we still need to determine its value of data λ. To do this, we can take the first-order condition with respect to Q:

$$\int (p_{e,i} + \lambda\alpha_i)\, \beta_i f_i(\Delta_i) di - \lambda = 0,$$

and then use the ecosystem's pricing expression to derive

$$\lambda = \int \beta_i F_i(\Delta_i) di. \tag{9}$$

To understand this expression, notice that if the ecosystem generates an extra unit of data, the first-order benefit is that it can raise price by β_i in market i and keep the same demand as before, thus earning an extra $\beta_i F_i(\Delta_i)$ on that product.

Since all the prices and quality investments characterized above depend on $\{\Delta_i\}$, we still need to determine $\{\Delta_i\}$ in order to fully solve the equilibrium. Using the above expressions for prices and qualities, we can rewrite the definition of Δ_i in (2) as

$$\Delta_i = 2\bar{\eta}_i F_i(\Delta_i) + \frac{1 - 2F_i(\Delta_i)}{f_i(\Delta_i)} - \eta_{s,i} + \alpha_i \lambda + \beta_i Q, \tag{10}$$

where $\bar{\eta}_i = (\eta_{e,i} + \eta_{s,i})/2$ is the average innovation efficiency in market i. Since both λ and Q are functions of $\{\Delta_i\}$, this yields a system of equations for $\{\Delta_i\}$. To solve this system, we solve the two aggregators λ and Q first. Under the conditions that we will specify below, the right-hand side of (10) has a slope of less than one in Δ_i, and so (10) uniquely determines Δ_i as a continuous function of λ and Q, which we denote by $\Delta_i(\lambda, Q)$. Substituting it into the expression for λ in (9) and the definition of Q in (3), we derive a system of equations in λ and Q:

$$\begin{aligned} \lambda &= \int \beta_i F_i(\Delta_i(\lambda, Q))di, \\ Q &= \int \alpha_i F_i(\Delta_i(\lambda, Q))di. \end{aligned} \tag{11}$$

Brouwer's fixed point theorem implies that this system has a solution with $\lambda \in [0, \mathbb{E}[\beta]]$ and $Q \in [0, \mathbb{E}[\alpha]]$, where $\mathbb{E}[\beta] \equiv \int \beta_i di$ and $\mathbb{E}[\alpha] \equiv \int \alpha_i di$ (as earlier). Once we solve λ and Q, we can determine Δ_i from (10) and then all the prices and qualities.

3.3 Equilibrium Existence and Uniqueness

Existence of the interior equilibrium characterized so far requires that (i) equation (10) has a solution $\Delta_i \in (-l_i, l_i)$ for each $i \in \mathcal{I}$, and (ii) no firms have profitable unilateral global deviations. In addition, it is helpful for the subsequent comparative statics analysis if there is a unique interior equilibrium. We now derive conditions for the existence and uniqueness of such an interior equilibrium.

For convenience, define

$$\sigma_i(x) \equiv 1 + \frac{d}{dx}\left(\frac{1 - F_i(x)}{f_i(x)}\right) = -\frac{[1 - F_i(x)]f_i'(x)}{f_i(x)^2}, \tag{12}$$

and

$$\chi \equiv \mathbb{E}[\alpha\beta] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}. \tag{13}$$

Note that $\sigma_i(x)$ measures the curvature of $1 - F_i(x)$ and $\sigma_i(x) \leq 1$ under our log-concavity assumption, while χ captures the strength of the ecosystem's data spillovers.

In the subsequent analysis, we make the following assumption on primitives:

Assumption 1. *For any $i \in \mathcal{I}$,*

$$\bar{\eta}_i + \frac{\chi}{2} < \frac{3}{2f_i(0)}, \tag{A1}$$

and

$$\max\{\eta_{s,i}, \eta_{e,i} + \alpha_i\mathbb{E}[\beta] + \beta_i\mathbb{E}[\alpha]\} < \frac{3}{2f_i(0)}, \tag{A2}$$

and

$$\max\{\eta_{s,i}, \eta_{e,i} + \chi\} < \min_{-l_i \leq x \leq l_i} \frac{2 - \sigma_i(x)}{f_i(x)}. \tag{A3}$$

Assumption 1 holds provided innovation efficiencies and data spillovers are small relative to the amount of product differentiation, as captured by $f_i(0)$ being sufficiently low on each product.[22] We then find that:

Proposition 1. *Under Assumption 1, there exists a unique interior equilibrium, and it is characterized by* (7)*,* (8)*,* (10)*, and* (11)*.*

(All omitted proofs can be found in the Appendix.) In the proof, we first use the Gale-Nikaido Theorem to show that (A1) implies that the mapping in (11) is injective (i.e., one-to-one) and therefore has a unique solution. We also show that (A1) implies that the right-hand side of (10) has a slope of less than one in Δ_i, so the solution Δ_i is unique (and stable). Second, we show that (A2) ensures that this unique solution is interior, that is, $\Delta_i \in (-l_i, l_i)$ in each market i. Finally, we prove that (A3) implies that no firm has a unilateral global deviation, and so the first-order conditions are sufficient in defining the equilibrium. This part is the most challenging given the ecosystem's optimization problem is of infinite dimension. As mentioned before, we deal with this issue by working with the system of demand quantity and quality.[23]

[22] To see that (A3) holds when $f_i(0)$ is sufficiently small, notice that log-concavity of f_i implies $\sigma_i(x) \leq 1$ and its symmetry implies $f_i(0) \geq f_i(x)$, and hence $\frac{1}{f_i(0)} \leq \frac{2-\sigma_i(x)}{f_i(x)}$ for any $x \in [-l_i, l_i]$.

[23] Condition (A3) also implies the condition in Lemma 1 for the consumer participation game. To see this, note that (A3) implies $\chi < 2/f_i(0)$ because $\sigma_i(0) = 0$; meanwhile, the Cauchy-Schwarz inequality implies $\sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]} \geq \mathbb{E}[\alpha\beta]$, and so $\chi \geq 2\mathbb{E}[\alpha\beta]$. Therefore, $\mathbb{E}[\alpha\beta] < 1/f_i(0)$ for each $i \in \mathcal{I}$.

We now provide some intuition for the role of conditions (A1)-(A3). Suppose, contrary to these conditions, that innovation efficiencies $\eta_{e,i}, \eta_{s,i}$ and/or data spillovers are large relative to product differentiation. First, multiple equilibria might exist: if the ecosystem, say, is expected to have higher sales, it invests much more and offers a much better product due to data spillovers, while small firms invest much less and so offer a much worse product. With low product differentiation, the ecosystem is able to generate enough extra sales to rationalize the initial expectations of these sales being higher. Second, these self-fulfilling expectations could even lead to corner equilibria where, in some markets, only one firm makes strictly positive sales. Thirdly, global deviations may be profitable: if the ecosystem, say, offers much lower prices on some products, due to strong data spillovers the demand for all its products could go up by so much that its profit increases.

A solvable example: the linear Hotelling case. Suppose that in each market i the distribution function is $F_i(x) = \frac{1}{2} + \frac{x}{2l_i}$, where l_i captures the degree of product differentiation in that market.[24] Then prices and investments in market i are

$$p_{e,i} = 2l_i z_i - \alpha_i \lambda,\ v_{e,i} = \eta_{e,i} z_i \quad \text{and} \quad p_{s,i} = 2l_i(1-z_i),\ v_{s,i} = \eta_{s,i}(1-z_i),$$

and the ecosystem's output in market i is

$$z_i = r_i + \frac{\alpha_i \lambda + \beta_i Q}{g_i},$$

where $g_i = 2(3l_i - \bar{\eta}_i)$, and where $r_i = (3l_i - \eta_{s,i})/g_i$ denotes the ecosystem's sales in the case where there are no data spillovers. Since each $z_i \equiv F_i(\Delta_i)$ is a linear function of λ and Q, the system in (11) is also linear, and hence can be solved explicitly as follows:

$$\begin{bmatrix} 1 - \mathbb{E}\left[\alpha\beta/g\right] & -\mathbb{E}[\beta^2/g] \\ -\mathbb{E}[\alpha^2/g] & 1 - \mathbb{E}[\alpha\beta/g] \end{bmatrix} \begin{bmatrix} \lambda \\ Q \end{bmatrix} = \begin{bmatrix} \mathbb{E}[r\beta] \\ \mathbb{E}[r\alpha] \end{bmatrix}.$$

We then obtain closed form solutions for equilibrium outputs, prices, and investments.

[24] In this example Assumption 1 simplifies to: $\bar{\eta}_i + \frac{\chi}{2} < 3l_i$, $\max\{\eta_{s,i}, \eta_{e,i} + \alpha_i\mathbb{E}[\beta] + \beta_i\mathbb{E}[\alpha]\} < 3l_i$, and $\max\{\eta_{s,i}, \eta_{e,i} + \chi\} < 4l_i$.

3.4 Cross-Market Comparison

We now investigate how prices and investments vary across markets with different data spillovers. All else equal, one might expect the ecosystem to charge less in markets that are better at generating data—but it turns out this is not always true.

Consider two markets j and k. Suppose they differ in data spillovers but are otherwise identical, that is, $F_j = F_k = F$, $\eta_{e,j} = \eta_{e,k} = \eta_e$ and $\eta_{s,j} = \eta_{s,k} = \eta_s$.

Proposition 2. *Suppose that two markets j and k differ only in data spillovers.*
(i) The ecosystem sells and invests more in market j than market k, while single-product firm j sells, invests and charges less than single-product firm k, if and only if

$$\alpha_j \lambda + \beta_j Q > \alpha_k \lambda + \beta_k Q. \tag{14}$$

(ii) The ecosystem can charge more in market j than in market k even if $\alpha_j > \alpha_k$.

Start with part (i) of the proposition. When comparing markets j and k we should treat (λ, Q) as fixed because there is a continuum of markets. Therefore equation (10) implies that if markets j and k differ only in their data spillovers, the ecosystem sells more in market j than in market k (i.e., $\Delta_j > \Delta_k$) if and only if (14) holds. (To interpret this condition, note that $\alpha_i \lambda + \beta_i Q$ represents product i's data-driven contribution to the ecosystem.) Since investment is proportional to demand, the investment results follow as well. Intuitively, in markets with high α, the ecosystem invests and sells more to exploit their greater ability to generate data; in markets with high β, the ecosystem sells more due to its quality advantage from data spillovers, which also makes investment more worthwhile.

Now consider the pricing results in the proposition. Using equations (7) and (8), the prices charged in market i by the single-product firm and the ecosystem are

$$p_{s,i} = \frac{1 - F(\Delta_i)}{f(\Delta_i)} \quad \text{and} \quad p_{e,i} = \frac{F(\Delta_i)}{f(\Delta_i)} - \lambda \alpha_i.$$

Given our log-concavity condition, single-product firm j charges strictly less than single-product firm k if and only if $\Delta_j > \Delta_k$. The comparison of ecosystem prices, however, is more subtle. On the one hand, if the ecosystem sells more in a market, this is a force towards a higher price via the $F(\Delta_i)/f(\Delta_i)$ term in the $p_{e,i}$ expression. On the other hand, though, if that same market is also better at generating data, this is a force towards a lower price via the subsidy term $-\lambda\alpha_i$. If markets j and k differ only in how well they use data (i.e., if $\alpha_j = \alpha_k$, but $\beta_j \neq \beta_k$), only the first effect is present, and the ecosystem charges more in whichever market uses data the best. Otherwise it is a priori ambiguous whether the ecosystem charges more in market j or market k.

We prove part (ii) of the proposition by constructing examples. In the linear Hotelling example solved before, one can check that

$$p_{e,j} > p_{e,k} \quad \Longleftrightarrow \quad (\alpha_j - \alpha_k)\lambda(\bar{\eta} - 2l) + (\beta_j - \beta_k)Ql > 0. \tag{15}$$

Now consider two interesting special cases. First, suppose the products are equally good at using data, but product j is better at generating data (i.e., $\alpha_j > \alpha_k$ but $\beta_j = \beta_k$). Then the ecosystem sells more of product j, and from (15) it also charges more on product j provided $\bar{\eta} > 2l$. Second, suppose product j only generates data while product k only uses data (i.e., $\alpha_j > \beta_j = 0$ and $\beta_k > \alpha_k = 0$). Then, from (15), the ecosystem can charge more on product j if $\bar{\eta} > 2l$, and also if $\alpha_j\lambda > \beta_k Q$ (i.e., if it sells more of product j than product k).[25] To understand these seemingly counterintuitive results, note that in both cases the ecosystem sells more of product j. As a result, the marginal "cost" of raising j's sales by cutting its price is higher. Hence, if η_e is large, it is relatively cheap to raise demand for j through investment, allowing the ecosystem to also raise its price for j. This is compounded when η_s is large—because single-product firm j reduces its investment strongly in anticipation of having lower demand. This in turn makes it even cheaper for the ecosystem to boost demand solely by raising investment, again allowing it to charge more in market j.

4 Data Regulation

In this section we use our framework to evaluate the effect of two regulatory policies: one that restricts the ecosystem's use of data, and another that forces the ecosystem to share data with small firms. We show that although each policy benefits small firms, they do not always benefit consumers, and indeed can sometimes help the ecosystem.

[25] To see the last point, note from Assumption 1 that $\bar{\eta} < 3l$ is required in this case, so $\alpha_j\lambda > \beta_k Q$ is a necessary condition (along with $\bar{\eta} > 2l$) for the condition in (15) to hold.

4.1 Restricting Data Usage

Consider a data policy that restricts data usage across at least part of the ecosystem's businesses: precisely, consider a policy which strictly reduces β_i in a positive measure of markets. (We note that privacy policies, which reduce α_i in a positive measure of markets, have qualitatively the same effect.) We find that:

Proposition 3. *(i) Restricting the ecosystem's data usage in a subset of markets induces it to sell less and innovate less in all markets, and induces all single-product firms to sell more, innovate more, set higher prices, and earn more profit.*
(ii) It benefits consumers if small firms are sufficiently efficient in innovation, i.e., if

$$\eta_{s,i} > \frac{1 - \sigma_i(\Delta_i) + F_i(\Delta_i)}{f_i(\Delta_i)} \tag{16}$$

in each market $i \in \mathcal{I}$, and harms consumers if the opposite is true in each market $i \in \mathcal{I}$.

Intuitively, restricting data usage (even only in a subset of markets) makes data less valuable for the ecosystem (i.e., λ decreases). In equilibrium this induces the ecosystem to sell less (i.e., Δ_i decreases) and innovate less in *all* markets, including those not directly affected by the regulation, and collect less data (i.e., Q decreases). Since the regulation shifts demand towards the single-product firms, they optimally raise their prices and invest more. (This is easy to see from (7) given f_i is log-concave.) Despite their higher innovation expenditure, the fact they sell more at a higher price means that the regulation increases their profit. To see this last point formally, note that using our earlier expressions for prices and qualities, the profit of single-product firm i can be written as

$$\Pi_{s,i} = [1 - F_i(\Delta_i)]^2 \left[\frac{1}{f_i(\Delta_i)} - \frac{\eta_{s,i}}{2} \right], \tag{17}$$

which given Assumption 1 is decreasing in Δ_i (and hence is increasing in single-product firm i's output).[26]

The effect of data regulation on the ecosystem's price and profit is more subtle. Specifically, recall from earlier that on product i the ecosystem charges

$$p_{e,i} = \underbrace{F_i(\Delta_i)/f_i(\Delta_i)}_{(-)} - \underbrace{\lambda\alpha_i}_{(-)}.$$

[26] Using the definition of $\sigma_i(x)$ in (12) one can check that $\Pi_{s,i}$ is decreasing in Δ_i provided $\eta_{s,i} < \frac{2-\sigma_i(\Delta_i)}{f_i(\Delta_i)}$, which is implied by (A3) in Assumption 1.

Data regulation reduces the quality of the ecosystem's product, which leads to a lower price (first term) given f_i is log-concave, but it also reduces the value of data λ and so reduces the ecosystem's incentive to subsidize data collection, which leads to a higher price (second term). In general either effect can dominate. Meanwhile, using our earlier expressions for prices and qualities, the ecosystem's profit can be written as

$$\Pi_e = \int \underbrace{F_i(\Delta_i)^2 \left[\frac{1}{f_i(\Delta_i)} - \frac{\eta_{e,i}}{2}\right]}_{(-)} di - \underbrace{\lambda Q}_{(-)}. \tag{18}$$

Data regulation causes the ecosystem to sell less of each product, which is a force towards lower profit (first term). Formally, given Assumption 1, the integrand is increasing in Δ_i.[27] However data regulation also reduces the value and volume of data collected by the ecosystem, and so reduces the total subsidy $\lambda Q \left(= \int \lambda \alpha_i F_i(\Delta_i) di\right)$ offered by the ecosystem, which is a force towards higher profit (second term). In general either effect can dominate as we will show in examples below.[28]

Regulation also has an ambiguous impact on consumers. To see this, first note that consumer surplus in market i is $V_i = \mathbb{E}[\max\{v_{e,i} - p_{e,i} + \beta_i Q + \epsilon_{e,i}, v_{s,i} - p_{s,i} + \epsilon_{s,i}\}]$. Using $\Delta_i = v_{e,i} - p_{e,i} + \beta_i Q - (v_{s,i} - p_{s,i})$, we can rewrite it as

$$V_i - \mathbb{E}[\max\{\epsilon_{e,i}, \epsilon_{s,i}\}] = v_{s,i} - p_{s,i} + \int_0^{\Delta_i} F_i(x) dx. \tag{19}$$

(This expression is valid regardless of the sign of Δ_i.) Hence, up to a constant, consumer surplus aggregated across all markets V can be expressed as

$$V = \int \left[\underbrace{v_{s,i}}_{(+)} - \underbrace{p_{s,i}}_{(+)} + \underbrace{\int_0^{\Delta_i} F_i(x) dx}_{(-)}\right] di. \tag{20}$$

Data regulation benefits consumers by boosting small firms' investment (first term), but harms consumers by inducing small firms to raise their prices (second term). Since data regulation causes a shift in demand away from the ecosystem, it also reduces the relative

[27] As in the last footnote, using the definition of $\sigma_i(x)$ and the symmetry of $f_i(x)$ one can check that this term is increasing in Δ_i provided $\eta_{e,i} < \frac{2-\sigma_i(-\Delta_i)}{f_i(-\Delta_i)}$, which is again implied by (A3) in Assumption 1.

[28] One may wonder why the ecosystem does not voluntarily reduce data usage if, say, it benefits from regulation that reduces some β_i. One reason may be a lack of commitment power: even though ex ante the ecosystem would like to use less data, once all prices and investments have been chosen, it should exploit data as much as it can so as to maximize consumer demand for its products.

quality of the ecosystem's products, which again harms consumers (third term). Notice that if $\eta_{s,i} = 0$ in each $i \in \mathcal{I}$, then small firms' innovation is completely unresponsive to the regulation—the first term is therefore irrelevant, and so regulation is unambiguously *bad* for consumers. (The same observation applies if $\eta_{e,i} = \eta_{s,i} = 0$ in each market, in which case our model reduces to one of pure price competition.) Therefore, for data regulation to benefit consumers, it must induce enough additional quality investment from small firms; this can only happen when small firms are sufficiently efficient in innovation, as indicated by condition (16).[29]

In the remainder of this section, we use two special cases to further explore the impact on data regulation on ecosystem profit and consumer welfare. We start with the case of symmetric products, and then consider the case of small data spillovers.

4.1.1 Symmetric products

Suppose that for each $i \in \mathcal{I}$: $F_i = F$ (so that $l_i = l$), $\eta_{e,i} = \eta_e$, $\eta_{s,i} = \eta_s$, $\alpha_i = 1$, and $\beta_i = \beta$. (In this symmetric case, α and β play the same role, so we normalize the former to 1.) Then Assumption 1 simplifies to

$$\max\{\eta_s, \eta_e + 2\beta\} < \min_{-l \le x \le l} \left\{ \frac{3}{2f(0)}, \frac{2-\sigma(x)}{f(x)} \right\}. \tag{21}$$

Under this condition, the equilibrium prices and investments in each market are

$$p_e = \frac{F(\Delta)}{f(\Delta)} - \beta F(\Delta),\ v_e = \eta_e F(\Delta) \quad \text{and} \quad p_s = \frac{1-F(\Delta)}{f(\Delta)},\ v_s = \eta_s[1-F(\Delta)]$$

where $\Delta \in (-l, l)$ uniquely solves

$$\Delta = 2AF(\Delta) + \left(\frac{1}{f(\Delta)} - \eta_s\right)(1 - 2F(\Delta)) \tag{22}$$

with

$$A \equiv \beta + \overline{\eta} - \eta_s$$

where $\overline{\eta} = (\eta_e + \eta_s)/2$.[30] Here A indicates the ecosystem's overall "advantage," including the strength of its data advantage β and its relative efficiency in investment $\overline{\eta} - \eta_s =$

[29]Note that (16) is not a primitive condition, but later we report examples where it is satisfied.

[30]Equation (22) is derived from (10) by using $\lambda = \beta F(\Delta)$ and $Q = F(\Delta)$. To see it has a unique solution, notice that condition (21) implies $\max\{\eta_s, \eta_e + 2\beta\} < \frac{3}{2f(0)}$ and so $\overline{\eta} + \beta < \frac{3}{2f(0)}$. One can check that this ensures the right-hand side of (22) has a slope of less than one in Δ.

$(\eta_e - \eta_s)/2$. It is easy to see that when $A = 0$, we have $\Delta = 0$, so the two firms split each market equally; one can also show that the ecosystem sells more and earns a higher profit in each market than the small firm if and only if $A > 0$.

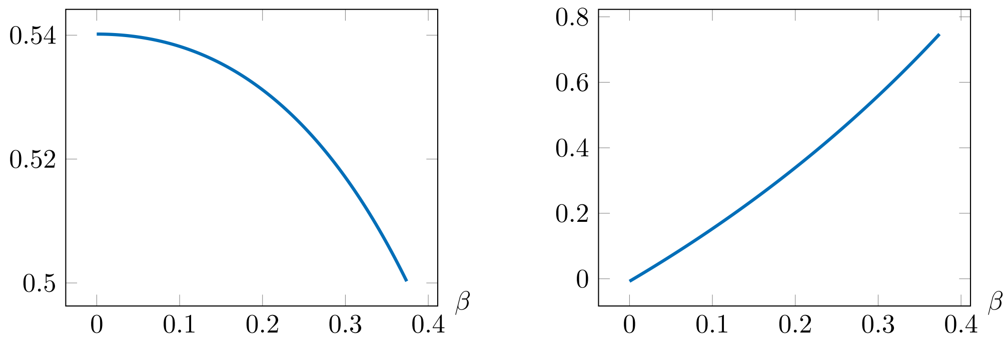
Proposition 4. *Suppose all the products are symmetric.*
(i) Starting from the "symmetric" situation where $A = 0$, data regulation that restricts cross-market data usage (i.e., reduces β) harms both the ecosystem and consumers.
(ii) In general, data regulation can benefit consumers or the ecosystem, but it cannot benefit both of them (i.e., it cannot lead to a Pareto improvement).

Result (i) shows that when markets are split equally, regulation harms both the ecosystem—by reducing its market share, and consumers—by worsening the ecosystem's surplus offer. Result (ii) shows that outside this symmetric case, regulation can potentially benefit consumers or the ecosystem, but not both at the same time. For instance, in order for consumers to benefit, the regulation must induce a sufficiently large increase in single-product firms' investment, but in that case the fall in ecosystem demand is so large that the ecosystem is made worse off.

The linear Hotelling example. When F is a uniform distribution on $[-l, l]$, we can further show that data regulation harms the ecosystem if and only if $A < l$ and harms consumers if $A \geq 0$.[31] Therefore, if $A > l$ (i.e., if the ecosystem initially has a sufficiently large advantage), interestingly data regulation *benefits* the ecosystem and at the same time *harms* consumers, which might be contrary to the original intention of the regulation. This is demonstrated in Figure 2 where the ecosystem is much more efficient at innovation: data regulation induces only a relatively small increase in single-product firms' innovation but sufficiently softens price competition, so that consumers are made worse off. In contrast, if $A < 0$ (i.e., if the small firms initially have an advantage), data regulation can benefit consumers. This is demonstrated in Figure 3 where small firms are much more efficient at innovation. Regulation that lowers β induces a relatively large increase in single-product firms' innovation; as a result regulation unambiguously harms the ecosystem, and benefits consumers provided β is not too large. Intuitively, consumers gain due to higher single-product firm innovation, but lose because the data-

[31] In this example, $\sigma(\Delta) = 0$ and so condition (21) simplifies further to $\max\{\eta_s, \eta_e + 2\beta\} < 3l$. One can also derive $\Delta = Al/(3l - A - \eta_s)$ from (22). A detailed proof of the claim is available upon request.

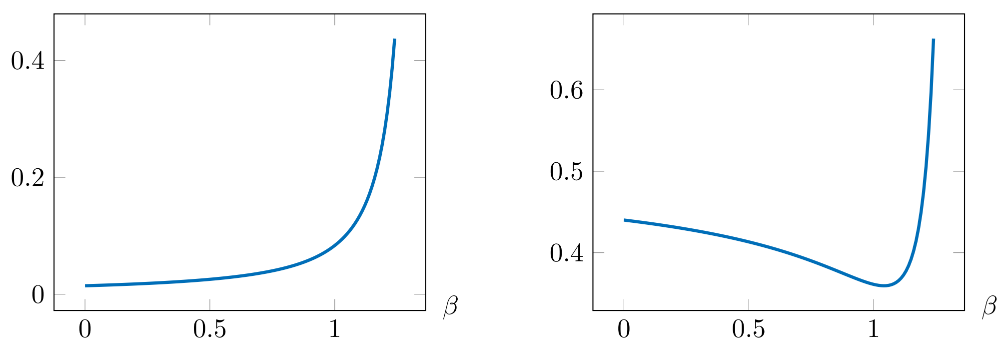
augmented part βQ of ecosystem surplus falls; provided β is not too large, the former effect dominates.



(a) Ecosystem Profit

(b) Consumer Surplus

Figure 2: The impact of restricing cross-market data usage
(products are symmetric, F is uniform on $[-1, 1]$, and $\alpha = 1, \eta_e = 2.25, \eta_s = 0.25$)



(a) Ecosystem Profit

(b) Consumer Surplus

Figure 3: The impact of restricting cross-market data usage
(products are symmetric, F is uniform on $[-1, 1]$, and $\alpha = 1, \eta_e = 0.5, \eta_s = 2.75$)

Breaking up the ecosystem. Our setup can also be used to discuss the impact of breaking up the ecosystem. Suppose the measure of all products is $N = \int_{\mathcal{I}'} di$ instead of 1. In the general setup with heterogeneous products, our analysis remains unchanged provided we replace expressions of the form $\mathbb{E}[x_i]$ with $\int_{\mathcal{I}'} x_i di$. In the setup of this section with symmetric products, our analysis also remains unchanged as long as we replace β by $N\beta$.[32] Then if, for example, we break up the ecosystem into two units

[32]More specifically, the ecosystem's price becomes $p_e = \frac{F(\Delta)}{f(\Delta)}[1 - N\beta f(\Delta)]$, and the ecosystem's advantage in equation (22) becomes $A = N\beta + \bar{\eta} - \eta_s$.

of equal size, the equilibrium outcome in each market will be the same as when the ecosystem has a measure $N/2$ of products. This effect, however, is the same as halving β. In other words, in our model data regulation can achieve the same effect as breaking up the ecosystem. In particular, breaking up the ecosystem dampens its incentive to invest, but induces small firms to invest more. Whether consumers benefit from the ecosystem being broken up depends on how responsive is small firms' investment.

4.1.2 Small data spillovers

Another case where we can make further analytical progress is when data spillovers are small. Suppose that the data spillover parameters are $\{\varepsilon\beta_i\}$ with $\varepsilon \approx 0$. Let r_i denote the ecosystem's demand in market i when there are *zero* data spillovers: $r_i = F_i(\Delta_i)$ where Δ_i solves (10) with $\lambda = 0$ and $\beta_i = 0$. Notice that $r_i > 1/2$ if and only if $\eta_{e,i} > \eta_{s,i}$ (i.e., if the ecosystem is more efficient in innovation in market i). As detailed in the appendix, we can derive the first-order Taylor approximation of the equilibrium variables and then show the following results:

Proposition 5. *Suppose the data spillovers are small (i.e., $\{\varepsilon\beta_i\}$ with $\varepsilon \approx 0$).*
(i) Data regulation that restricts cross-market data usage harms the ecosystem if

$$\eta_{e,i} - \eta_{s,i} < \frac{1}{f_i} - \frac{f_i'}{f_i^3} \tag{23}$$

for each i, where $f_i = f_i(F_i^{-1}(r_i))$ and $f_i' = f_i'(F_i^{-1}(r_i))$, and benefits the ecosystems if the opposite holds for each i.
(ii) It benefits consumers if

$$\eta_{s,i} > \frac{1 - \sigma_i(r_i) + r_i}{f_i} \tag{24}$$

for each i, where $\sigma_i(r_i) = -(1 - r_i)f_i'/f_i^2$, and harms consumers if the opposite holds for each i.
(iii) It cannot benefit both the ecosystem and consumers (i.e., it cannot lead to a Pareto improvement) if $F_i = F$, $\eta_{e,i} = \eta_e$, and $\eta_{s,i} = \eta_s$ for each i. With product heterogeneity, however, data regulation can benefit all firms and consumers.

If $\eta_{e,i} = \eta_{s,i}$, then $r_i = 1/2$ and $f_i' = 0$, so (23) must hold. In other words, if the ecosystem and small firms are equally efficient in innovation in every market, data regulation harms the ecosystem when data spillovers are small. In the linear Hotelling example, (23) simplifies to $\eta_{e,i} - \eta_{s,i} < 2l_i$. Given $\eta_{e,i} < 3l_i$ under Assumption 1, this condition must hold if $\eta_{s,i} > \min\{l_i, \eta_{e,i}\}$, i.e., if each small firm is sufficiently efficient in innovation. Condition (24) is the counterpart of (16) when data spillovers are small. In the linear Hotelling example with $\sigma_i(r_i) = 0$, this condition simplifies to $\eta_{s,i} > 2l_i(1+r_i)$ where $r_i = \frac{3l_i - \eta_{s,i}}{2(3l_i - \bar{\eta}_i)}$. This condition holds if and only if $\eta_{s,i}$ exceeds a threshold less than $3l_i$ since r_i decreases in $\eta_{s,i}$ and $r_i \to 0$ as $\eta_{s,i} \to 3l_i$.

In Section 4.1.1 we have shown that if all products are symmetric (including their data effects), it is impossible for data regulation to lead to a Pareto improvement (i.e., either the ecosystem or consumers will get harmed). With small data spillovers, the same result holds without requiring symmetry of data effects. More interestingly, when products are heterogeneous, there exist examples where data regulation leads to a Pareto improvement, i.e., where for the ecosystem the reduction of data subsidies caused by data regulation dominates while for consumers the increased innovation by small firms dominates.

4.2 Data Sharing

Let us now consider an alternative policy that mandates the ecosystem to share its data with single-product firms. If a consumer buys single-product firm i's product, we assume that she now obtains surplus

$$v_{s,i} - p_{s,i} + \gamma_i Q + \epsilon_{s,i},$$

where $\gamma_i \geq 0$ measures how effectively single-product firm i can use the shared data to improve its product quality. The basic surplus difference between the two firms in market i then becomes

$$\Delta_i = v_{e,i} - p_{e,i} + (\beta_i - \gamma_i)Q - (v_{s,i} - p_{s,i}).$$

Compared to the baseline model, the difference is that now the ecosystem's *net* data advantage $\beta_i - \gamma_i$ matters.

Suppose $\beta_i > \gamma_i$ in each market, reflecting the fact that either the ecosystem is better at using data due to, say, its technology advantage, or it only shares part of its data with single-product firms. Once we replace β_i in the baseline model by the net data advantage $\beta_i - \gamma_i$, the equilibrium analysis remains unchanged. Then the equilibrium existence and uniqueness result in Proposition 1 carries over, and we also have:

Proposition 6. *(i) Data sharing that increases γ_i in a positive measure of markets has the same impact on prices, investments, sales and profits as restricting data usage that decreases β_i in the same set of markets.*
(ii) Data sharing is more beneficial/less harmful to consumers than restricting data usage in the ecosystem.

To understand part (i), notice that decreasing β_i or increasing γ_i has the same effect on the ecosystem's net data advantage $\beta_i - \gamma_i$, and hence has the same effect on outputs, prices, investments, and profits. However, as shown by part (ii), the two policies do not have the same effect on consumer surplus. To see this, notice that under data sharing consumer surplus in market i is $V_i = \mathbb{E}[\max\{v_{e,i} - p_{e,i} + \beta_i Q + \epsilon_{e,i}, v_{s,i} - p_{s,i} + \gamma_i Q + \epsilon_{s,i}\}]$, which can be written as

$$V_i - \mathbb{E}[\max\{\epsilon_{e,i}, \epsilon_{s,i}\}] = v_{s,i} - p_{s,i} + \gamma_i Q + \int_0^{\Delta_i} F_i(x)dx. \tag{25}$$

Compared to the baseline model, this has an extra term $\gamma_i Q$ due to single-product firms' quality improvement from shared data. Data sharing has the same effect on $v_{s,i} - p_{s,i}$, Q and Δ_i as restricting data usage, but it has an extra positive effect via the $\gamma_i Q$ term. In general, however, data sharing also has an ambiguous impact on consumer surplus since it reduces the ecosystem's product quality and can raise all market prices.

A joint policy. Finally, it is easy to see from the above analysis that if we raise β_i and γ_i by the same amount in a positive measure of markets, the utility offered by the ecosystem relative to the small firms remains unchanged—and hence it has no effect on Δ_i in any market. Therefore, this joint policy has no impact on equilibrium prices, investments, and firm profits. However consumer surplus must increase, because the data part of utility is higher at both the ecosystem ($\beta_i Q$) and the small firms ($\gamma_i Q$). Therefore, allowing the ecosystem to use more data (e.g., when it obtains more data or its data technology improves), but also forcing it to share more data with smaller firms—in such a way that $\beta_i - \gamma_i$ remains unchanged—is unambiguously good for consumers.

5 Data cooperative

In this section we consider the possibility that single-product firms form a data cooperative to share their data with each other (but their price and innovation decisions remain independent). We show that such a data cooperative does not necessarily benefit single-product firms, and it benefits consumers when products are sufficiently symmetric to each other, but can otherwise harm consumers.

We first extend the baseline model by introducing a data cooperative. If a consumer buys the ecosystem's product i, she obtains the same surplus

$$v_{e,i} - p_{e,i} + \beta_{e,i} Q_e + \epsilon_{e,i}$$

as before, where we now use $\beta_{e,i}$ to indicate the product specific data-spillover effects within the ecosystem, and where

$$Q_e = \int \alpha_i F_i(\Delta_i) di$$

is the amount of data the ecosystem possesses; if the consumer buys from single-product firm i, she now obtains a surplus

$$v_{s,i} - p_{s,i} + \beta_{s,i} Q_s + \epsilon_{s,i},$$

where

$$Q_s = \int \alpha_i [1 - F_i(\Delta_i)] di$$

is the amount of data the cooperative possesses, and the new quality improvement term $\beta_{s,i} Q_s$ is from data sharing among single-product firms themselves. We allow for different extents of data spillover effects between the ecosystem and the data cooperative.

Since $Q_e + Q_s = \mathbb{E}[\alpha]$,[33] in the following we let $Q_e = Q$ and $Q_s = \mathbb{E}[\alpha] - Q$. Then

$$\Delta_i = v_{e,i} - p_{e,i} + (\beta_{e,i} + \beta_{s,i}) Q - (v_{s,i} - p_{s,i} + \beta_{s,i} \mathbb{E}[\alpha]). \tag{26}$$

Compared to the baseline model, each firm's demand is now more sensitive to how much data the ecosystem has.

Notice that, with a continuum of products, no single-product firm's price and quality choices can affect the amount of data the cooperative possesses. Therefore, this extended model can be solved by following the same logic as in the baseline model. To ensure the existence and uniqueness of an (interior) equilibrium, Assumption 1 needs to be strengthened to the following:

[33] This identity relies on the implicit assumption that how much data a product generates does not depend on who sells it.

Assumption 2. *For any $i \in \mathcal{I}$,*

$$\bar{\eta}_i + \frac{\chi}{2} < \frac{3}{2f_i(0)}, \tag{B1}$$

and

$$\max\{\eta_{s,i}, \eta_{e,i} + \alpha_i \mathbb{E}[(\beta_e + \beta_s)] + (\beta_{e,i} + \beta_{s,i})\mathbb{E}[\alpha]\} + \beta_{s,i}\mathbb{E}[\alpha] < \frac{3}{2f_i(0)}, \tag{B2}$$

and

$$\max\{\eta_{s,i}, \eta_{e,i} + \chi\} < \min_{-l_i \le x \le l_i} \frac{2 - \sigma_i(x)}{f_i(x)}, \tag{B3}$$

where

$$\chi = \mathbb{E}[\alpha(\beta_e + \beta_s)] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[(\beta_e + \beta_s)^2]}.$$

When $\beta_{s,i} = 0$ for every $i \in \mathcal{I}$, this assumption degenerates to Assumption 1. With $\beta_{s,i} > 0$ for a positive measure of i, χ is greater than in the baseline case, so all the conditions are more stringent than in Assumption 1. Intuitively, this is because the overall data-spillover effect is now stronger: when the ecosystem has more data, single-product firms will have less, and so the ecosystem's relative data advantage is more sensitive to Q as we have seen from (26). As before, this new assumption must hold if $f_i(0)$ is sufficiently small in each market.

The equilibrium characterization of price and quality investment is the same as in the baseline model, except that now we have

$$p_{e,i} = \frac{F_i(\Delta_i)}{f_i(\Delta_i)} - \alpha_i \lambda \quad \text{with} \quad \lambda = \int (\beta_{e,i} + \beta_{s,i}) F_i(\Delta_i) dj.$$

That is, for fixed $\{\Delta_i\}$, the ecosystem now offers larger data subsidies than before. This is because when the ecosystem lowers its prices, this also reduces each single-product firm's "quality" (through the term $\beta_{s,i} Q_s = \beta_{s,i}(\mathbb{E}[\alpha] - Q)$) and is hence more profitable than in the baseline case. This suggests, as we will show below, that single-product firms do not necessarily benefit from forming a data cooperative, as doing so will trigger more aggressive pricing by the ecosystem.

The equation for Δ_i becomes

$$\Delta_i = 2\bar{\eta}_i F_i(\Delta_i) + \frac{1 - 2F_i(\Delta_i)}{f_i(\Delta_i)} - \eta_{s,i} + \alpha_i \lambda + (\beta_{e,i} + \beta_{s,i})Q - \beta_{s,i}\mathbb{E}[\alpha]. \tag{27}$$

Since the right-hand side has a slope of less than one in Δ_i under Assumption 2, we can solve Δ_i as a function of (λ, Q) and then derive a system of equations:

$$\lambda = \int (\beta_{e,i} + \beta_{s,i}) F_i(\Delta_i(\lambda, Q)) di \quad \text{and} \quad Q = \int \alpha_i F_i(\Delta_i(\lambda, Q)) di. \tag{28}$$

This system has a unique solution with $\lambda \in [0, \mathbb{E}[\beta_e + \beta_s]]$ and $Q \in [0, \mathbb{E}[\alpha]]$ under condition (B1).

Since $\Pi_{s,i}$ takes the same form as in the baseline model, it must decrease in Δ_i. From the Δ_i equation, we can see that if $\beta_{s,i}$ alone increases while other $\beta_{s,j\neq i}$'s remain unchanged, then the right-hand side of (27) will be smaller, as the change of one firm's data effect does not affect (λ, Q). Therefore, Δ_i will decrease and so single-product firm i will earn a higher profit.

When a positive measure of $\beta_{s,i}$'s increase, however, the impact on single-product firms is subtler as (λ, Q) will change as well. For example, if both λ and Q increase, there is a counter force for Δ_i to increase. As we will see below, strengthening the data cooperative can indeed boost the ecosystem's sales and *harm* single-product firms.

5.1 Symmetric products

To make more progress, let us first consider the case with symmetric products, and as before we normalize $\alpha = 1$. Then the subsidy term in the ecosystem's price becomes $\lambda = (\beta_e + \beta_s)F(\Delta)$, and the Δ equation simplifies to

$$\Delta = 2AF(\Delta) + \left(\frac{1}{f(\Delta)} - \eta_s - \beta_s\right)(1 - 2F(\Delta)), \tag{29}$$

where $A = \beta_e + \overline{\eta} - \eta_s$ is the ecosystem's overall (dis)advantage, the same as defined in the baseline model and independent of β_s. Note that $\Delta = 0$ if $A = 0$. Hence, given the right-hand side of (29) has a slope of less than one in Δ under Assumption 2, we have $\Delta > 0$ if and only if $A > 0$, the same as in the baseline case. Compared to (22), the only difference here is the extra $-\beta_s(1 - 2F(\Delta))$ term on the right-hand side. It is positive if and only if $\Delta > 0$. This immediately implies the following result:

Lemma 2. *If $A > 0$, (29) has a unique solution $\Delta > 0$ and it increases in β_s; if $A < 0$, (29) has a unique solution $\Delta < 0$ and it decreases in β_s. Therefore, introducing the data cooperative will boost single-product firms' sales and lower the ecosystem's sales if and only if $A < 0$ (i.e., if single-product firms already have an advantage in the market).*

Having the data cooperative improves each single-product firm's quality by $\beta_s[1-F(\Delta)]$, but it also induces the ecosystem to lower its price by $\beta_s F(\Delta)$, i.e, the extra term in its data subsidy. The former effect dominates if and only if $\Delta < 0$ or $A < 0$.

Aggregate consumer surplus is (up to a constant)

$$V = v_s - p_s + \beta_s[1 - F(\Delta)] + \int_0^\Delta F(x)dx. \tag{30}$$

If Δ were the same as in the baseline case (in which case $v_s - p_s$ would be the same as well), having a data cooperative would improve consumer welfare by increasing the overall quality of single-product firms. With an endogenous Δ, the impact is less obvious; however, with symmetric products we can show that consumers always benefit from introducing the data cooperative.

Proposition 7. *Suppose all products are symmetric. A data cooperative among single-product firms benefits its members if and only if $A < 0$ in the baseline model (i.e., if they initially have an overall advantage), harms the ecosystem if $A < \hat{A}$ for some $\hat{A} > 0$, and always benefits consumers.*

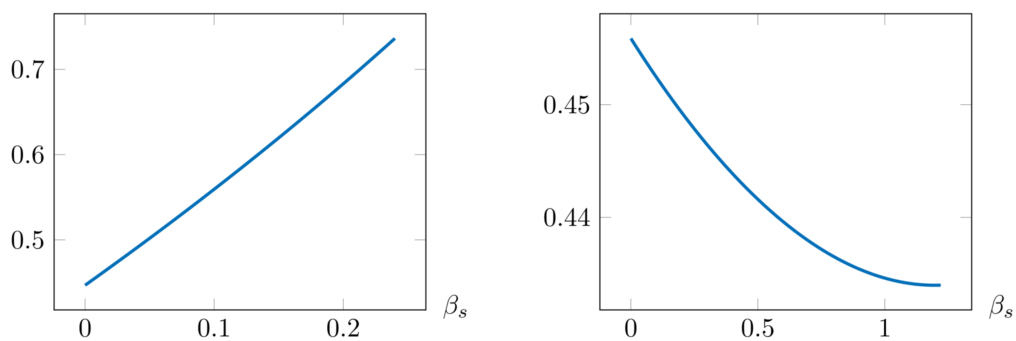
The result concerning single-product firms follows immediately from Lemma 2, as their profit decreases in Δ. The result concerning the ecosystem implies that whenever single-product firms have an incentive to form a data cooperative, this must harm the ecosystem by inducing it to sell less. (It also implies that at least for some small $A > 0$, the data cooperative harms *both* small firms and the ecosystem.) When the ecosystem sells less, it also innovates less and this can harm consumers. However, the consumer result indicates that this potential negative effect on consumers is outweighed by the positive effects: small firms' quality improves due to the data-spillover effect, and the ecosystem prices more aggressively. (As we will see below, however, this result relies on the assumption of symmetric products.) The consumer result also opens the door for another policy that would subsidize small firms to establish a data cooperative when they are in a disadvantageous position in the market.

5.2 Small data spillovers

Now consider the case with potentially asymmetric products but small data spillovers $\{\varepsilon\beta_{e,i}, \varepsilon\beta_{s,i}\}$ where $\varepsilon \approx 0$. We can obtain the following additional observations:

Proposition 8. *Suppose the data spillovers are small (i.e., $\{\varepsilon\beta_{e,i}, \varepsilon\beta_{s,i}\}$ where $\varepsilon \approx 0$).*
(i) If introducing the data cooperative induces the same extent of data spillovers across two single-product firms, the one that is better at generating data (i.e., has a higher α_i) is more likely to suffer.
(ii) With asymmetric products, introducing the data cooperative can harm consumers.

Intuitively, in responding to the data cooperative, the ecosystem will price more aggressively in those data-rich markets, and so the small firms there are more likely to suffer. This explains result (i). This result also suggests that small firms in those data-rich markets have less incentive to form a data cooperative, while other small firms may have incentive to subsidize them. Introducing the data cooperative can induce the ecosystem to innovate less, and this negative quality effect can dominate under product heterogeneity. In the proof, we construct such an example where in a small set of data-rich markets small firms are sufficiently efficient in innovation, whereas in other markets with limited data generation the ecosystem is sufficiently efficient.



(a) Symmetric markets

(b) Asymmetric markets

Figure 4: The impact of a data cooperative on consumer surplus
(F is uniform on $[-1, 1]$. In the left panel markets are symmetric, with $\alpha = 1$, $\beta_e = 0.05$, $\eta_e = 0$, and $\eta_s = 2.75$. In the right panel 10% of markets are as in the left panel, while the remaining 90% of markets have $\alpha = 0$, $\beta_e = 0.05$, $\eta_e = 2.75$, and $\eta_s = 0$, and β_s is the same in all markets.)

Figure 4 depicts the effect of increasing β_s, and hence strengthening a data cooperative, on consumers. (Note that $\beta_s = 0$ corresponds to the case with no cooperative.) In the left panel all products are symmetric, and consumer surplus monotonically increases in β_s as predicted in Proposition 7. In the right panel markets are asymmetric and, consistent with the above results on small spillovers, consumer surplus decreases in β_s (which is assumed to be common across markets).

6 Conclusion

This paper makes two contributions. First, it provides a framework to study competition between a multiproduct digital ecosystem and many single-product rivals. The framework features data spillovers across markets, allows for both price and innovation competition, and accommodates rich product heterogeneity. To the best of our knowledge, such a framework is new to the literature. Second, the paper uses this framework to evaluate various data policies, such as restricting cross-market data usage, mandating data sharing from the ecosystem to small competitors, and facilitating exchange of data among small firms via a data cooperative. The first two policies always benefit small firms and encourage them to innovate more, but do not necessarily benefit consumers because they reduce ecosystem investment and may cause all prices in the market to increase. The third policy of creating a data cooperative can actually harm the small firms, because it induces more aggressive pricing from the ecosystem; it benefits consumers when markets are sufficiently symmetric to each other but otherwise can harm consumers due to a reduction in the ecosystem's innovation.

The framework presented in this paper could be further developed and used to address many other interesting questions related to digital ecosystems. For instance, instead of focusing on policies that regulate the use of data, we could examine policies that affect market structure, such as allowing small firms to merge and become larger competitors. Moreover, in this paper we have exogenously fixed the ecosystem's product range, but it would be interesting to endogenize its choice of which products to supply. We have also focused on the case with one large ecosystem; further research could explore a more general case with multiple ecosystems, which may have only partially overlapping businesses, competing both with each other and with single-product firms. Finally, to focus on cross-product data usage, we have adopted a static model, though in practice data is accumulated and updated over time. Developing a fully dynamic model in our context can be challenging; however, our equilibrium could represent the steady state when the ecosystem's data decays over time while new consumers enter and contribute fresh data. We plan to investigate some of these issues in future work.

References

ACCC (2023): *Digital platform services inquiry, Interim report 7: Report on expanding ecosystems of digital platform service providers*. Australian Competition and Consumer Commission.

Anderson, S., and O. Bedre Defolie (2025): "App Platform Model," mimeo.

Anderson, S. P., A. de Palma, and J.-F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. The MIT Press.

Armstrong, M. (2006): "Competition in Two-Sided Markets," *The RAND Journal of Economics*, 37(3), 668–691.

Belleflamme, P., M. Peitz, and E. Toulemonde (2022): "The tension between market shares and profit under platform competition," *International Journal of Industrial Organization*, 81, 102807.

Beraja, M., D. Y. Yang, and N. Yuchtman (2023): "Data-intensive Innovation and the State: Evidence from AI Firms in China," *The Review of Economic Studies*, 90(4), 1701–1723.

Bisceglia, M., and J. Tirole (2024): "Fair Gatekeeping in Digital Ecosystems," mimeo.

Chen, Z., and P. Rey (2023): "A Theory of Conglomerate Mergers," mimeo.

Condorelli, D., and J. Padilla (2024): "Data-Driven Envelopment with Privacy-Policy Tying," *The Economic Journal*, 134(658), 515–537.

Fan, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Review*, 103(5), 1598–1628.

Farboodi, M., R. Mihet, T. Philippon, and L. Veldkamp (2019): "Big Data and Firm Dynamics," *AEA Papers and Proceedings*, 109, 38–42.

Hagiu, A., and J. Wright (2023): "Data-enabled learning, network effects, and competitive advantage," *The RAND Journal of Economics*, 54(4), 638–667.

He, Z., J. Huang, and J. Zhou (2023): "Open banking: Credit market competition when borrowers own the data," *Journal of Financial Economics*, 147(2), 449–474.

Heidhues, P., M. Köster, and B. Kőszegi (2024): "A Theory of Digital Ecosystems," mimeo.

Jeon, D.-S., Y. Lefouili, Y. Li, and T. Simcoe (2024): "Ecosystems and Complementary Platforms," mimeo.

Jones, C. I., and C. Tonetti (2020): "Nonrivalry and the Economics of Data," *American Economic Review*, 110(9), 2819–58.

Jullien, B., A. Pavan, and M. Rysman (2021): "Two-Sided Markets, Pricing, and Network effects," in *Handbook of Industrial Organization, Volume 4*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, vol. 4, pp. 485–592. Elsevier.

Katz, M. L., and C. Shapiro (1985): "Network Externalities, Competition, and Compatibility," *The American Economic Review*, 75(3), 424–440.

Kraemer, J., and S. Shekhar (2024): "Regulating Digital Platform Ecosystems through Data-Sharing and Data-Siloing: Consequences for Innovation and Welfare," *MIS Quarterly forthcoming*.

Nalebuff, B. (2004): "Bundling as an Entry Barrier," *The Quarterly Journal of Economics*, 119(1), 159–187.

Peitz, M., and S. Sato (2023): "Asymmetric Platform Oligopoly," mimeo.

Peukert, C., A. Sen, and J. Claussen (2024): "The Editor and the Algorithm: Recommendation Technology in Online News," *Management Science*, 70(9), 5816–5831.

Prüfer, J., and C. Schottmüller (2021): "Competing with Big Data," *The Journal of Industrial Economics*, 69(4), 967–1008.

Rhodes, A., and J. Zhou (2019): "Consumer Search and Retail Market Structure," *Management Science*, 65(6), 2607–2623.

Rochet, J.-C., and J. Tirole (2006): "Two-Sided Markets: A Progress Report," *The RAND Journal of Economics*, 37(3), 645–667.

Rysman, M. (2004): "Competition Between Networks: A Study of the Market for Yellow Pages," *The Review of Economic Studies*, 71(2), 483–512.

Schaefer, M., and G. Sapi (2023): "Complementarities in learning from data: Insights from general search," *Information Economics and Policy*, 65, 101063.

Tan, G., and J. Zhou (2021): "The Effects of Competition and Entry in Multi-sided Markets," *The Review of Economic Studies*, 88(2), 1002–1030.

Whinston, M. (1990): "Tying, Foreclosure, and Exclusion," *American Economic Review*, 80(4), 837–859.

Yoganarasimhan, H. (2020): "Search Personalization Using Machine Learning," *Management Science*, 66(3), 1045–1070.

Zhou, J. (2025): "Firewall for Innovation," mimeo.

A Appendix: Omitted Proofs

A.1 Proof of Proposition 1

First, we show that condition (A1) implies that (11) has a unique solution (λ, Q). To have an explicit expression for $\Delta_i(\lambda, Q)$, let us define

$$G_i(x) \equiv F_i^{-1}(x) - 2\bar{\eta}_i x - \frac{1-2x}{f_i(F_i^{-1}(x))}. \tag{31}$$

Then the Δ_i equation (10) can be rewritten as $G_i(F_i(\Delta_i)) = -\eta_{s,i} + \alpha_i\lambda + \beta_i Q$. As shown below, condition (A1) implies that G_i is an increasing function. We can then derive

$$F_i(\Delta_i) = G_i^{-1}(-\eta_{s,i} + \alpha_i\lambda + \beta_i Q), \tag{32}$$

and rewrite (11) as

$$\begin{aligned} \lambda - \int \beta_i G_i^{-1}(-\eta_{s,i} + \alpha_i\lambda + \beta_i Q)di = 0, \\ Q - \int \alpha_i G_i^{-1}(-\eta_{s,i} + \alpha_i\lambda + \beta_i Q)di = 0. \end{aligned} \tag{33}$$

Brouwer's fixed point theorem implies that this system must have a solution. We now use the Gale-Nikaido Theorem to show that the solution is unique.

The mapping defined in the left-hand side of (33) has a rectangular domain $[0, \mathbb{E}[\beta]] \times [0, \mathbb{E}[\alpha]]$. Its Jacobian matrix is

$$\begin{bmatrix} 1 - \int \frac{\alpha_i\beta_i}{g_i}di & -\int \frac{\beta_i^2}{g_i}di \\ -\int \frac{\alpha_i^2}{g_i}di & 1 - \int \frac{\alpha_i\beta_i}{g_i}di \end{bmatrix} = \begin{bmatrix} 1-C & -B \\ -A & 1-C \end{bmatrix}$$

where $g_i = g_i(G_i^{-1}(\cdot))$ and $g_i(\cdot) = G_i'(\cdot)$. Note that from $G_i(x)$ defined in (31), we have

$$\begin{aligned} g_i(x) &= \frac{3}{f_i(F_i^{-1}(x))} - 2\bar{\eta}_i + \frac{(1-2x)f_i'(F_i^{-1}(x))}{f_i(F_i^{-1}(x))^3} \\ &\geq \frac{3}{f_i(F_i^{-1}(x))} - 2\bar{\eta}_i \geq \frac{3}{f_i(0)} - 2\bar{\eta}_i > \chi, \end{aligned} \tag{34}$$

where the first inequality is because the symmetry and log-concavity of f_i imply that $f_i'(F_i^{-1}(x)) > 0$ for $x < 1/2$ and $f_i'(F_i^{-1}(x)) < 0$ for $x > 1/2$, the second inequality is because $f_i(0) \geq f_i(F_i^{-1}(x))$, and the last one is from condition (A1). (Given $\chi > 0$, $G_i(x)$ is indeed monotonic as claimed before.) Using this observation, we have

$$0 < C < C + \sqrt{AB} < \frac{\mathbb{E}[\alpha\beta] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}}{\chi} = 1. \tag{35}$$

Therefore, *at any point*, all the principal minors of the Jacobian have a strictly positive determinant:

$$1 - C > 0, \quad (1 - C)^2 - AB > 0.$$

That is, the Jacobian is a P-matrix and so the mapping is injective (i.e., one-to-one). This implies uniqueness of the solution (λ, Q).

Second, we show that (10) has an interior solution $\Delta_i \in (-l_i, l_i)$ under condition (A2). To have an interior solution, it is sufficient to establish that

$$-l_i < \frac{1}{f_i(-l_i)} - \eta_{s,i} \quad \text{and} \quad l_i > \eta_{e,i} - \frac{1}{f_i(l_i)} + \alpha_i \mathbb{E}[\beta] + \beta_i \mathbb{E}[\alpha]$$

given $0 \leq \lambda \leq \mathbb{E}[\beta]$ and $0 \leq Q \leq \mathbb{E}[\alpha]$. Using $f_i(-l_i) = f_i(l_i)$, we can rewrite these two conditions as

$$\max\{\eta_{s,i}, \eta_{e,i} + \alpha_i \mathbb{E}[\beta] + \beta_i \mathbb{E}[\alpha]\} < l_i + \frac{1}{f_i(l_i)}.$$

Given the log-concavity and symmetry of f_i, we have $f_i(l_i) \leq f_i(0)$ and $1 = \int_{-l_i}^{l_i} f_i(x)dx \leq \int_{-l_i}^{l_i} f_i(0)dx = 2l_i f_i(0)$, and so $l_i + \frac{1}{f_i(l_i)} \geq \frac{1}{2f_i(0)} + \frac{1}{f_i(0)} = \frac{3}{2f_i(0)}$. Therefore a sufficient condition for the above inequality is condition (A2).[34]

Third, we show that (10) has a unique (stable) solution under condition (A1). It suffices to prove that the right-hand side of (10) has a slope of less than one in Δ_i, which has already been shown to be true under condition (A1).

Finally, we show that given other firms adopt their equilibrium strategies, no firm has a profitable unilateral (global) deviation. This issue is easier to deal with when we reformulate each firm's optimization problem as a system of quality and demand quantity (instead of price). Recall that $z_i = F_i(\Delta_i)$ denotes the ecosystem's demand in market i, where

$$\Delta_i = v_{e,i} - p_{e,i} + \beta_i Q - v_{s,i} + p_{s,i}.$$

For single-product firm i, from $z_i = F_i(\kappa_i - v_{s,i} + p_{s,i})$ where $\kappa_i = v_{e,i} - p_{e,i} + \beta_i Q$, we have $p_{s,i} = F_i^{-1}(z_i) - \kappa_i + v_{s,i}$. Then its optimization problem, given the ecosystem's $\{v_{e,i}, p_{e,i}\}_{i \in \mathcal{I}}$, can be rewritten as

$$\max_{z_i, v_{s,i}} \; [F_i^{-1}(z_i) - \kappa_i + v_{s,i}](1 - z_i) - \frac{v_{s,i}^2}{2\eta_{s,i}}. \tag{36}$$

[34] Note that (A2) is tight in the uniform-distribution case, but otherwise there is some slack. In particular, if $l_i = \infty$ no conditions are needed for an interior solution. However we use (A2) in some later results.

(Note that there is a one-to-one mapping between $(p_{s,i}, v_{s,i})$ and $(z_i, v_{s,i})$.) The objective function is clearly concave in $v_{s,i}$, so the optimal $v_{s,i}$ must be equal to $\eta_{s,i}(1-z_i)$. Then single-product firm i's problem simplifies to

$$\max_{z_i} \ \Pi_{s,i}(z_i) \equiv [F_i^{-1}(z_i) - \kappa_i](1-z_i) + \frac{\eta_{s,i}}{2}(1-z_i)^2. \tag{37}$$

It is straightforward to verify that the first-order condition here yields the same characterization as in the price-quality approach. The first-order condition is also sufficient if $\Pi_{s,i}(z_i)$ is strictly concave, or equivalently if $\eta_{s,i} < -[F_i^{-1}(z_i)(1-z_i)]''$ for any $z_i \in [0,1]$. Note that

$$-[F_i^{-1}(z_i)(1-z_i)]'' = \frac{2f_i(x)^2 + [1-F_i(x)]f_i'(x)}{f_i(x)^3} = \frac{2-\sigma_i(x)}{f_i(x)}, \text{ where } x = F_i^{-1}(z_i).$$

(Recall that under the log-concavity condition, $\sigma_i(x) \leq 1$, so this expression must be positive.) Therefore, single-product firm i has no profitable unilateral deviation if

$$\eta_{s,i} < \min_{x\in[-l_i,l_i]} \frac{2-\sigma_i(x)}{f_i(x)}.$$

Now consider the ecosystem's problem. From $z_i = F_i(v_{e,i} - p_{e,i} + \zeta_i)$ where $\zeta_i = \beta_i Q - v_{s,i} + p_{s,i}$, we have $p_{e,i} = v_{e,i} - F_i^{-1}(z_i) + \zeta_i$. Then the ecosystem's problem, given single-product firms' $\{v_{s,i}, p_{s,i}\}_{i\in\mathcal{I}}$, can be rewritten as

$$\max_{\mathbf{z},\mathbf{v_e}} \int \left([v_{e,i} - F_i^{-1}(z_i) + \zeta_i] z_i - \frac{v_{e,i}^2}{2\eta_{e,i}} \right) di \quad \text{s.t.} \quad Q = \int \alpha_i z_i di.$$

(There is also a one-to-one mapping between $(p_{e,i}, v_{e,i})$ and $(z_i, v_{e,i})$.) The integrand is clearly concave in $v_{e,i}$, so the optimal $v_{e,i}$ must be equal to $\eta_{e,i} z_i$. Then the problem simplifies to

$$\max_{\mathbf{z}} \ \Pi_e(\mathbf{z}) = \int \left(\frac{\eta_{e,i}}{2} z_i - F_i^{-1}(z_i) + \zeta_i \right) z_i di \quad \text{s.t.} \quad Q = \int \alpha_i z_i di.$$

Substituting $\zeta_i = \beta_i Q - v_{s,i} + p_{s,i}$ and the constraint into the objective function, the problem becomes

$$\max_{\mathbf{z}} \ \Pi_e(\mathbf{z}) = \int \left(\frac{\eta_{e,i}}{2} z_i - F_i^{-1}(z_i) - v_{s,i} + p_{s,i} \right) z_i di + \int \alpha_i z_i di \int \beta_i z_i di.$$

We claim that $\Pi_e(\mathbf{z})$ is concave in $\mathbf{z}$ if

$$\eta_{e,i} + \chi < \min_{z_i\in[0,1]} [z_i F_i^{-1}(z_i)]'' = \min_{x\in[-l_i,l_i]} \frac{2-\sigma_i(x)}{f_i(x)} \tag{38}$$

for any i, where $\chi = \mathbb{E}[\alpha\beta] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}$. The equality is because

$$[z_i F_i^{-1}(z_i)]'' = \frac{2f_i(x)^2 - F_i(x)f_i'(x)}{f_i(x)^3} = \frac{2f_i(-x)^2 + [1 - F_i(-x)]f_i'(-x)}{f_i(-x)^3} = \frac{2 - \sigma_i(-x)}{f_i(-x)},$$

where $x = F_i^{-1}(z_i)$ and the second equality uses the symmetry of f_i, and

$$\min_{-l_i \le -x \le l_i} \frac{2 - \sigma_i(-x)}{f_i(-x)} = \min_{-l_i \le x \le l_i} \frac{2 - \sigma_i(x)}{f_i(x)}.$$

To prove the above claim, notice that under (38), we can always find ξ_i satisfying

$$\chi < \xi_i < \min[z_i F_i^{-1}(z_i)]'' - \eta_{e,i}. \tag{39}$$

Rewrite $\Pi_e(\mathbf{z})$ as the sum of two terms:

$$\Pi_e(\mathbf{z}) = \int \left(\frac{\eta_{e,i} + \xi_i}{2} z_i - F_i^{-1}(z_i) - v_{s,i} + p_{s,i} \right) z_i di + \left(\int \alpha_i z_i di \int \beta_i z_i di - \frac{1}{2} \int \xi_i z_i^2 di \right)$$

The first term is concave in $\mathbf{z}$ because, for each i, the integrand is strictly concave in z_i by observing that

$$\left[\left(\frac{\eta_{e,i} + \xi_i}{2} z_i - F_i^{-1}(z_i) - v_{s,i} + p_{s,i} \right) z_i \right]'' = \eta_{e,i} + \xi_i - [z_i F_i^{-1}(z_i)]'' < 0,$$

where the inequality uses (39).

The second term in $\Pi_e(\mathbf{z})$ is a quadratic form of $\mathbf{z}$. It is also concave if it is negative semidefinite in the sense that it is negative for *any* $\mathbf{z}$ (even if some z_i's are negative).[35] To show that, we first prove the following inequality:

$$\int \alpha_i z_i di \int \beta_i z_i di \le \frac{\chi}{2} \int z_i^2 di. \tag{40}$$

Notice that by the Cauchy-Schwarz inequality we have

$$\left(\int \left(\alpha_i \sqrt{\mathbb{E}[\beta^2]} + \beta_i \sqrt{\mathbb{E}[\alpha^2]} \right)^2 di \right) \left(\int z_i^2 di \right) \ge \left(\int \left(\alpha_i \sqrt{\mathbb{E}[\beta^2]} + \beta_i \sqrt{\mathbb{E}[\alpha^2]} \right) z_i di \right)^2.$$

[35] Formally, suppose $H(\mathbf{z})$ is a quadratic form of $\mathbf{z}$. Then, for any $\lambda \in (0,1)$, $\forall \mathbf{z}', \mathbf{z}''$,

$$\lambda H(\mathbf{z}') + (1-\lambda)H(\mathbf{z}'') - H(\lambda \mathbf{z}' + (1-\lambda)\mathbf{z}'') = \lambda(1-\lambda)H(\mathbf{z}' - \mathbf{z}'').$$

If the quadratic form is negative semidefinite in the sense that $H(\mathbf{z}) \le 0$ for any $\mathbf{z}$, then $\lambda H(\mathbf{z}') + (1-\lambda)H(\mathbf{z}'') - H(\lambda \mathbf{z}' + (1-\lambda)\mathbf{z}'') \le 0$, which implies concavity of H.

The left-hand side is equal to

$$2\left(\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]+\mathbb{E}[\alpha\beta]\sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}\right)\int z_i^2 di.$$

The right-hand side is equal to

$$\left(\sqrt{\mathbb{E}[\beta^2]}\int\alpha_i z_i di+\sqrt{\mathbb{E}[\alpha^2]}\int\beta_i z_i di\right)^2\geq 4\sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}\int_0^1\alpha_i z_i di\int_0^1\beta_i z_i di$$

where the inequality uses the arithmetic mean-geometric mean (AM-GM) inequality, i.e., $(A+B)^2\geq 4AB$. Combining these observations proves (40). Then it immediately follows that

$$\int\alpha_i z_i di\int\beta_i z_i di-\frac{1}{2}\int\xi_i z_i^2 di\leq\int\frac{\chi-\xi_i}{2}z_i^2 di\leq 0$$

for any $\mathbf{z}$, where the first inequality uses (40) and the second uses (39). Therefore, the second term in $\Pi_e(\mathbf{z})$ is concave as well.

In sum, all firms' profit functions are concave in their own quantity choices if

$$\max\{\eta_{s,i},\eta_{e,i}+\chi\}<\min_{x\in[-l_i,l_i]}\frac{2-\sigma_i(x)}{f_i(x)}$$

for any i, which is condition (A3). This completes the whole proof. □

A.2 Proof of Proposition 3

(i) In this proof, for each variable x, we denote by $\dot{x}$ its corresponding marginal change due to the data regulation policy. Consider a policy that marginally changes data parameters (α_i,β_i) by $(\dot{\alpha}_i,\dot{\beta}_i)$ for each $i\in\mathcal{I}$. (We allow for zero marginal change in a subset of parameters or markets. For completeness, our proof here also allows for a data policy which affects the α_i's as well.) Recall that $z_i=F_i(\Delta_i)$, $\lambda=\int\beta_i z_i di$, and $Q=\int\alpha_i z_i di$. Differentiating λ and Q yields:

$$\dot{\lambda}=\int\dot{\beta}_i z_i di+\int\beta_i\dot{z}_i di\quad\text{and}\quad\dot{Q}=\int\dot{\alpha}_i z_i di+\int\alpha_i\dot{z}_i di.$$

Recall from (32) that $z_i=G_i^{-1}(-\eta_{s,i}+\alpha_i\lambda+\beta_i Q)$, so we have

$$\dot{z}_i=\frac{\dot{\alpha}_i\lambda+\alpha_i\dot{\lambda}+\dot{\beta}_i Q+\beta_i\dot{Q}}{g_i},$$

where $g_i = g_i(G_i^{-1}(\cdot))$ is defined in (34). Substituting this into $\dot{\lambda}$ and $\dot{Q}$ yields

$$\begin{bmatrix} \dot{\lambda} \\ \dot{Q} \end{bmatrix} = \mathbf{G} \begin{bmatrix} \dot{\lambda} \\ \dot{Q} \end{bmatrix} + \mathbf{d},$$

where

$$\mathbf{G} \equiv \begin{bmatrix} \int \frac{\alpha_i \beta_i}{g_i} di & \int \frac{\beta_i^2}{g_i} di \\ \int \frac{\alpha_i^2}{g_i} di & \int \frac{\alpha_i \beta_i}{g_i} di \end{bmatrix}$$

and

$$\mathbf{d} = \begin{bmatrix} \int \left\{ \dot{\beta}_i z_i + \frac{\beta_i}{g_i}(\dot{\alpha}_i \lambda + \dot{\beta}_i Q) \right\} di \\ \int \left\{ \dot{\alpha}_i z_i + \frac{\alpha_i}{g_i}(\dot{\alpha}_i \lambda + \dot{\beta}_i Q) \right\} di \end{bmatrix}$$

We then have

$$\begin{bmatrix} \dot{\lambda} \\ \dot{Q} \end{bmatrix} = [\mathbf{I} - \mathbf{G}]^{-1} \mathbf{d} = (\mathbf{I} + \mathbf{G} + \mathbf{G}^2 + \cdots) \mathbf{d},$$

where $\mathbf{I}$ is the identity matrix of size 2. The infinite sum $\mathbf{I} + \mathbf{G} + \mathbf{G}^2 + \cdots$ converges as both eigenvalues of $\mathbf{G}$ are in the interval $(-1, 1)$ by condition (A1) in Assumption 1.[36] This also shows every entry of $[\mathbf{I} - \mathbf{G}]^{-1}$ is positive.

Suppose $\dot{\alpha}_i \leq 0$ and $\dot{\beta}_i \leq 0$ for every i. Then we have $\mathbf{d} \leq \mathbf{0}$, which further implies that $\dot{\lambda} \leq 0$ and $\dot{Q} \leq 0$. (If for a positive measure of products $\dot{\alpha}_i < 0$ or $\dot{\beta}_i < 0$, then, at least one entry of $\mathbf{d}$ is negative, which implies that $\dot{\lambda} < 0$ and $\dot{Q} < 0$.) Then, for each market i, we have

$$\dot{z}_i = \frac{\dot{\alpha}_i \lambda + \alpha_i \dot{\lambda} + \dot{\beta}_i Q + \beta_i \dot{Q}}{g_i} \leq 0,$$

which implies that $\dot{\Delta}_i \leq 0$. It then follows immediately that $\dot{v}_{s,i} = -\eta_{s,i} f_i(\Delta_i) \dot{\Delta}_i \geq 0$ and $\dot{v}_{e,i} = \eta_{e,i} f_i(\Delta_i) \dot{\Delta}_i \leq 0$, and also

$$\dot{p}_{s,i} = \left[\frac{1 - F_i(\Delta_i)}{f_i(\Delta_i)} \right]' \dot{\Delta}_i \geq 0 \quad \text{and} \quad \dot{\Pi}_{s,i} = -[1 - F_i(\Delta_i)][2 - \sigma_i(\Delta_i) - \eta_{s,i} f_i(\Delta_i)] \dot{\Delta}_i \geq 0,$$

where the first inequality uses log-concavity of $1 - F_i$ and the second inequality uses condition (A3) for the small firm's second-order condition.

[36]The eigenvalues are $C + \sqrt{AB}$ and $C - \sqrt{AB}$ where

$$A = \int \frac{\alpha_i^2}{g_i} di > 0, \; B = \int \frac{\beta_i^2}{g_i} di > 0, \; C = \int \frac{\alpha_i \beta_i}{g_i} di \geq 0.$$

As shown in equation (35) in the proof of Proposition 1, $0 < C + \sqrt{AB} < 1$. Meanwhile, $-1 < -(C + \sqrt{AB}) < C - \sqrt{AB} < C + \sqrt{AB} < 1$.

(ii) Let $V_i(\Delta_i) \equiv v_{s,i} - p_{s,i} + \int_0^{\Delta_i} F_i(x)dx$. If

$$V_i'(\Delta_i) = -\eta_{s,i} f_i(\Delta_i) + 1 - \sigma_i(\Delta_i) + F_i(\Delta_i) < 0,$$

or equivalently if (16) holds in each market, data regulation benefits consumers. When the opposite is true in each market, data regulation harms consumers. □

A.3 Proof of Proposition 4

(i) With symmetric products and the normalization $\alpha = 1$, the ecosystem's profit is

$$\Pi_e = F^2 \left(\frac{1}{f} - \frac{\eta_e}{2} \right) - \beta F^2,$$

where the second term is the data-induced subsidies offered by the ecosystem and is derived by using $\lambda = \beta F(\Delta)$ and $Q = F(\Delta)$. As explained in the general case, both terms increase in Δ so that the overall impact of data regulation on Π_e is ambiguous. One can check that[37]

$$\frac{d\Pi_e}{d\beta} = \frac{F}{2} \left(1 - 2Af - \frac{f'}{f^2} \right) \frac{d\Delta}{d\beta}, \tag{41}$$

where $A = \beta + \bar{\eta} - \eta_s$ is the ecosystem's advantage. Given Δ increases in β, this must be positive at the symmetric situation with $\Delta = 0$ (in which case $A = 0$ and $f' = 0$). That is, starting from the symmetric situation, data regulation that reduces β harms the ecosystem.

For consumer surplus, using the symmetric-product version of (19), one can check that

$$\frac{dV}{d\beta} = (-\eta_s f + 1 - \sigma + F) \frac{d\Delta}{d\beta}. \tag{42}$$

At the symmetric situation with $\Delta = 0$, it has the sign of $\frac{3}{2} - \eta_s f(0)$ given $\sigma(0) = 0$. This must be positive as condition (21) requires $\eta_s f(0) < \frac{3}{2}$. That is, starting from the symmetric situation, data regulation that reduces β harms consumers.

[37] We use

$$\frac{d\Delta}{d\beta} = \frac{2F}{3 - 2(\beta + \bar{\eta})f + (1 - 2F)f'/f^2} > 0,$$

which is derived from (22). Condition (21) ensures that the right-hand side of (22) has a slope of less than one in Δ, which is equivalent to the denominator above being positive.

(ii) We now show that it is impossible to have both $\frac{d\Pi_e}{d\beta} \leq 0$ and $\frac{dV}{d\beta} \leq 0$. From (41) and (42), we see that $\frac{d\Pi_e}{d\beta} \leq 0$ if and only if

$$\frac{f'}{f^2} \geq 1 - 2Af,$$

and $\frac{dV}{d\beta} \leq 0$ if and only if

$$\frac{f'}{f^2} \leq -\frac{1 + F - \eta_s f}{1 - F}.$$

In order for both of them to hold, we must have

$$1 - 2Af \leq -\frac{1 + F - \eta_s f}{1 - F} \iff \frac{2}{f} \leq 2A(1 - F) + \eta_s.$$

When $A \leq 0$, this is impossible since condition (21) requires $\eta_s \leq 3/(2f)$. When $A > 0$, this is also impossible since

$$2A(1 - F) + \eta_s \leq 2A + \eta_s = 2\beta + \eta_e < \frac{3}{2f},$$

where the equality used the definition of A and the last inequality used condition (21).

$\square$

A.4 Proof of Proposition 5

Let us first approximate the equilibrium prices and investments. Recall that r_i denotes the ecosystem's sales in market i when there are zero data spillovers. Define $\hat{\alpha} = \int \alpha_i r_i di$ and $\hat{\beta} = \int \beta_i r_i di$.

Using the G_i function defined in (31), we can write the system of (λ, Q) as

$$\lambda = \int \varepsilon \beta_i G_i^{-1}(-\eta_{s,i} + \alpha_i \lambda + \varepsilon \beta_i Q) di \quad \text{and} \quad Q = \int \alpha_i G_i^{-1}(-\eta_{s,i} + \alpha_i \lambda + \varepsilon \beta_i Q) di.$$

Note that $G_i^{-1}(-\eta_{s,i}) = r_i$. When $\varepsilon = 0$, we have $\lambda = 0$ and $Q = \int \alpha_i r_i di = \hat{\alpha}$. When $\varepsilon > 0$ is small, the first-order Taylor approximation of (λ, Q), after discarding higher-order terms, is

$$\lambda \approx \varepsilon \underbrace{\int \beta_i r_i di}_{\hat{\beta}} \quad \text{and} \quad Q \approx \hat{\alpha} + \varepsilon \int \alpha_i \underbrace{\frac{\alpha_i \hat{\beta} + \beta_i \hat{\alpha}}{g_i}}_{\theta_i} di \tag{43}$$

where

$$g_i = G_i'(r_i) = \frac{3}{f_i} - 2\bar{\eta}_i + \frac{(1-2r_i)f_i'}{f_i^3} = \frac{3 - \sigma_i(r_i) - \sigma_i(1-r_i)}{f_i} - 2\bar{\eta}_i > 0 \tag{44}$$

with $f_i = f_i(F_i^{-1}(r_i))$ and $f_i' = f_i'(F_i^{-1}(r_i))$, and the second equality used

$$\sigma_i(r_i) = -\frac{(1-r_i)f_i'}{f_i^2} \quad \text{and} \quad \sigma_i(1-r_i) = \frac{r_i f_i'}{f_i^2}$$

where the second expression will be proved below. Using these approximations, we derive

$$z_i = F_i(\Delta_i) = G_i^{-1}(-\eta_{s,i} + \alpha_i\lambda + \varepsilon\beta_i Q) \approx r_i + \theta_i\varepsilon. \tag{45}$$

Then we immediately have

$$v_{e,i} \approx \eta_{e,i}(r_i + \theta_i\varepsilon) \quad \text{and} \quad v_{s,i} \approx \eta_{s,i}(1 - r_i - \theta_i\varepsilon).$$

We can also approximate the ecosystem's prices as

$$p_{e,i} = \frac{z_i}{f_i(F_i^{-1}(z_i))} - \alpha_i\lambda \approx \frac{r_i}{f_i} + \left(\frac{1}{f_i} - \frac{r_i f_i'}{f_i^3}\right)\theta_i\varepsilon - \alpha_i\hat{\beta}\varepsilon.$$

Using the definition of $\sigma_i(r_i) = -\frac{(1-r_i)f_i'}{f_i^2}$, we can verify that

$$\sigma_i(1-r_i) = \frac{r_i f_i'(F_i^{-1}(r_i))}{f_i^2(F_i^{-1}(r_i))}$$

by using the symmetry of f_i and $F_i^{-1}(1-r_i) = -F_i^{-1}(r_i)$. Therefore, we have

$$p_{e,i} \approx \frac{r_i}{f_i} + \frac{1 - \sigma_i(1-r_i)}{f_i}\theta_i\varepsilon - \alpha_i\hat{\beta}\varepsilon.$$

Similarly, we can approximate small firm's price as

$$p_{s,i} = \frac{1 - z_i}{f_i(F_i^{-1}(z_i))} \approx \frac{1 - r_i}{f_i} - \frac{1 - \sigma_i(r_i)}{f_i}\theta_i\varepsilon.$$

The presence of data spillovers has an ambiguous effect on the ecosystem's price (as usual): it raises the ecosystem's quality, which leads to a higher price (second term in $p_{e,i}$), but also gives the ecosystem an incentive to collect data, which leads to a lower price (third term in $p_{e,i}$). Data spillovers unambiguously cause small firms to reduce prices due to reduction in their relative quality vis-à-vis the ecosystem (second term in $p_{s,i}$).

(i) The ecosystem's profit is

$$\Pi_e = \int \underbrace{z_i^2 \left[\frac{1}{f_i(F_i^{-1}(z_i))} - \frac{\eta_{e,i}}{2}\right]}_{\phi_i(z_i)} di - \lambda Q \approx \Pi_{e,0} + \left(\int P_i \theta_i di - \hat{\alpha}\hat{\beta}\right)\varepsilon,$$

where $\Pi_{e,0} = \int \phi_i(r_i)di$ is the ecosystem profit when there are no data spillovers, and

$$P_i = \phi_i'(r_i) = r_i \left(\frac{2 - \sigma_i(1 - r_i)}{f_i} - \eta_{e,i}\right).$$

(We have used $\lambda Q \approx \hat{\alpha}\hat{\beta}\varepsilon$ in the approximation.) Assumption 1 implies $P_i > 0$. Notice that we can rewrite the bracket term in the approximation as

$$\hat{\alpha} \int \left(\frac{P_i}{g_i} - \frac{r_i}{2}\right) \beta_i di + \hat{\beta} \int \left(\frac{P_i}{g_i} - \frac{r_i}{2}\right) \alpha_i di. \tag{46}$$

Therefore, reducing β_i in a positive measure of markets harms the ecosystem if $P_i/g_i > r_i/2$ for each i but benefits the ecosystem if $P_i/g_i < r_i/2$ for each i. One can check that

$$\frac{P_i}{g_i} > \frac{r_i}{2} \iff \frac{[2 - \sigma_i(1 - r_i)]/f_i - \eta_{e,i}}{[3 - \sigma_i(r_i) - \sigma_i(1 - r_i)]/f_i - 2\bar{\eta}_i} > \frac{1}{2} \iff \eta_{e,i} - \eta_{s,i} < \frac{1}{f_i} - \frac{f_i'}{f_i^3}.$$

(ii) Consumer surplus is (up to a constant)

$$V = \int \underbrace{\left[v_{s,i} - p_{s,i} + \int_0^{F_i^{-1}(z_i)} F_i(x)dx\right]}_{\varphi_i(z_i)} di \approx V_0 + \int S_i \theta_i di \times \varepsilon,$$

where $V_0 = \int \varphi_i(r_i)di$ is consumer surplus when there are no data spillovers, and

$$S_i = \varphi_i'(r_i) = \frac{1 + r_i - \sigma_i(r_i)}{f_i} - \eta_{s,i}.$$

It immediately follows that reducing β_i in a positive measure of markets harms consumers if $S_i > 0$ for each i and benefits consumers if $S_i < 0$ for each i. Note that

$$S_i > 0 \iff \eta_{s,i} < \frac{1 + r_i - \sigma_i(r_i)}{f_i}.$$

(iii) Suppose we reduce each β_i by τ_i (where τ_i can be zero in some markets). The approximation in (46) implies that this benefits the ecosystem if and only if

$$-\hat{\alpha} \int \left(\frac{P_i}{g_i} - \frac{r_i}{2}\right) \tau_i di - \int \tau_i r_i di \int \left(\frac{P_i}{g_i} - \frac{r_i}{2}\right) \alpha_i di > 0.$$

Using the definition of $\hat{\alpha}$, we can simplify it to

$$\hat{\alpha}\int\left(r_i-\frac{P_i}{g_i}\right)\tau_i di>\int r_i\tau_i di\int\frac{P_i}{g_i}\alpha_i di. \tag{47}$$

Similarly, data regulation benefits consumers if and only if

$$-\hat{\alpha}\int\frac{S_i}{g_i}\tau_i di>\int r_i\tau_i di\int\frac{S_i}{g_i}\alpha_i di. \tag{48}$$

When the product (but not necessarily the data) primitives are symmetric across products, (47) and (48) simplify to $P/g<r/2$ and $S<0$, respectively, given $g_i>0$. However, they cannot hold simultaneously because in the general case we have

$$\frac{P_i}{r_i}+S_i>g_i. \tag{49}$$

To see that, the left-hand side is equal to

$$\frac{2-\sigma_i(1-r_i)}{f_i}-\eta_{e,i}+\frac{1+r_i-\sigma_i(r_i)}{f_i}-\eta_{s,i}=g_i+\frac{r_i}{f_i}>g_i,$$

where the equality used (44).

On the other hand, when product primitives are asymmetric across products, it is possible that both (47) and (48) hold, and so data regulation leads to a Pareto improvement. To see that, suppose data primitives are symmetric across products and data regulation reduces spillover by τ in each market. Then (47) and (48) respectively simplify to

$$2\int\frac{P_i}{g_i}di<\int r_i di;\quad \int\frac{S_i}{g_i}<0.$$

(Note that it is impossible to have both $2P_i/g_i<r_i$ and $S_i<0$ for each i given $P_i>0$ and the inequality (49).) We now provide a linear Hotelling example where both conditions are satisfied. Suppose that half of the products are characterized by $(l_1,\eta_{e,1},\eta_{s,1})$ and the other half characterized by $(l_2,\eta_{e,2},\eta_{s,2})$. Suppose $\eta_{s,2}=3l_2>\eta_{e,2}$. (Our argument below holds when $\eta_{s,2}$ is sufficiently close to $3l_2$.) Then

$$r_2=\frac{3l_2-\eta_{s,2}}{2(3l_2-\bar{\eta}_2)}=0;\quad \frac{P_2}{g_2}=\frac{r_2(4l_2-\eta_{e,2})}{2(3l_2-\bar{\eta}_2)}=0;\quad \frac{S_2}{g_2}=\frac{2l_2(1+r_2)-\eta_{s,2}}{2(3l_2-\bar{\eta}_2)}=\frac{-l_2}{2(3l_2-\bar{\eta}_2)}.$$

The two conditions above simplify to

$$2\frac{P_1}{g_1}<r_1 \iff 2l_1<\eta_{e,1}-\eta_{s,1};\quad \frac{S_1}{g_1}-\frac{l_2}{2(3l_2-\bar{\eta}_2)}<0 \iff \frac{2l_1(1+r_1)-\eta_{s,1}}{3l_1-\bar{\eta}_1}<\frac{l_2}{3l_2-\bar{\eta}_2}.$$

Now suppose $\eta_{s,1} = 0$. Then the first condition is $2l_1 < \eta_{e,1}$, and a sufficient condition for the second one is

$$\frac{4l_1}{6l_1 - \eta_{e,1}} < \frac{l_2}{3l_2 - \eta_{e,2}}.$$

Given $\eta_{e,1} < 3l_1$, the left-hand side is no greater than $4/3$, so we can always pick an $\eta_{e,2}$ sufficiently close to $3l_2$ to make the above inequality hold. □

A.5 Proof of Proposition 7

A single-product firm's profit takes the same form as in the baseline model, so it decreases in Δ. Then Lemma 2 immediately implies the result concerning the profit of single-product firms.

Now consider the impact of data cooperative on the ecosystem's profit. Notice that

$$\Pi_e = F(\Delta)^2 \left(\frac{1}{f(\Delta)} - \frac{\eta_e}{2} - \beta_e - \beta_s \right) = F(\Delta)^2 \left(\frac{1}{f(\Delta)} - \frac{\eta_s}{2} - A - \beta_s \right).$$

(This differs from the baseline case by the β_s term.) Then

$$\frac{d\Pi_e}{d\beta_s} = -F^2 + F\left[2 - \frac{Ff'}{f^2} - 2\left(A + \beta_s + \frac{\eta_s}{2}\right) f\right] \frac{d\Delta}{d\beta_s}.$$

Using

$$\frac{d\Delta}{d\beta_s} = \frac{2F - 1}{3 - 2(A + \beta_s + \eta_s) f + (1 - 2F) f'/f^2}$$

(which is derived from (29) and where the denominator must be positive under condition (B1) in Assumption 2), one can check that

$$\begin{aligned}
\frac{d\Pi_e}{d\beta_s} &\propto F - 2 + 2(A + \beta_s)(1 - F) f + \eta_s f \\
&< F - 2 + \left(\frac{3}{2f(0)} - \eta_s \right) (1 - F) f + \eta_s f \\
&< F - 2 + \frac{3f}{2f(0)} \\
&\leq F - \frac{1}{2},
\end{aligned}$$

where the first inequality used $2(A + \beta_s) = 2(\beta_e + \beta_s) + \eta_e - \eta_s < \frac{3}{2f(0)} - \eta_s$ (which is implied by condition (B2) in Assumption 2), and the second inequality used $1 - F < 1$. Therefore, when $\Delta \leq 0$ and so $F \leq 1/2$, we must have $\frac{d\Pi_e}{d\beta_s} < 0$. Given Δ decreases in

β_s when it is negative, we can then claim that introducing a data cooperative among single-product firms will strictly harm the ecosystem if $\Delta \leq 0$ (or $A \leq 0$) in the baseline case. By continuity, the same result also holds at least when $\Delta > 0$ (or $A > 0$) is small.

Now consider consumer surplus. We have

$$\frac{dV}{d\beta_s} = \left[1 + F + \frac{(1-F)f'}{f^2} - (\beta_s + \eta_s)f\right]\frac{d\Delta}{d\beta_s} + 1 - F.$$

Substituting the above expression for $\frac{d\Delta}{d\beta_s}$, one can verify that

$$\frac{dV}{d\beta_s} \propto 2(1 - F + F^2) - (\beta_s + \eta_s)f - 2A(1-F)f.$$

When $\Delta < 0$, we have $A < 0$, and so the above expression must be greater than

$$2(1 - F + F^2) - (\beta_s + \eta_s)f > 0,$$

where the inequality used $2(1 - F + F^2) = 2(F - \frac{1}{2})^2 + \frac{3}{2} \geq \frac{3}{2}$ and $\beta_s + \eta_s < \frac{3}{2f(0)}$ (which is implied by Assumption 2). When $\Delta \geq 0$, we have $A \geq 0$ and $1 - F < \frac{1}{2}$. Then the above expression must be greater than

$$2(1 - F + F^2) - (\beta_s + \eta_s)f - Af = 2(1 - F + F^2) - (A + \beta_s + \eta_s)f > 0,$$

where the inequality used $2(1 - F + F^2) \geq \frac{3}{2}$ and $A + \beta_s + \eta_s = \bar{\eta} + \beta_e + \beta_s < \frac{3}{2f(0)}$ (which is also implied by Assumption 2). Therefore, $\frac{dV}{d\beta_s} > 0$ for all Δ, so consumers benefit from introducing the data cooperative. □

A.6 Proof of Proposition 8

Consider the case with asymmetric products but small data spillovers $\{\varepsilon\beta_{e,i}, \varepsilon\beta_{s,i}\}$ where $\varepsilon \approx 0$. Using the $G_i(\cdot)$ function defined in the baseline case, we can write the system of equations in (λ, Q) as

$$\lambda = \int (\beta_{e,i} + \beta_{s,i})F_i(\Delta_i)di \quad \text{and} \quad Q = \int \alpha_i F_i(\Delta_i)di,$$

with $F_i(\Delta_i) = G_i^{-1}(-\eta_{s,i} + \alpha_i\lambda + (\beta_{e,i} + \beta_{s,i})Q - \beta_{s,i}\mathbb{E}[\alpha])$. As before, let r_i be the ecosystem's sales in market i when there are no data spillovers. Following the same approximation procedure in Section 4.1.2, we can derive

$$\lambda \approx \varepsilon(\hat{\beta}_e + \hat{\beta}_s) \quad \text{and} \quad Q \approx \hat{\alpha} + \varepsilon \int \alpha_i \underbrace{\frac{\alpha_i(\hat{\beta}_e + \hat{\beta}_s) + (\beta_{e,i} + \beta_{s,i})\hat{\alpha} - \beta_{s,i}\mathbb{E}[\alpha]}{g_i}}_{\theta_i} di$$

where we used the notation $\hat{x} = \int r_i x_i di$. Then

$$z_i = G_i^{-1}(-\eta_{s,i} + \alpha_i \lambda + (\beta_{e,i} + \beta_{s,i})Q - \beta_{s,i}\mathbb{E}[\alpha]) \approx r_i + \theta_i \varepsilon. \tag{50}$$

Single-product firm i's profit $\Pi_{s,i}$ takes the same form as in the baseline case, so it decreases in z_i. Therefore, data cooperative improves $\Pi_{s,i}$ if and only if it decreases θ_i. Note that increasing $\{\beta_{s,i}\}$ only affects the numerator of θ_i which can be written as

$$\alpha_i \int r_j(\beta_{e,j} + \beta_{s,j})dj + \beta_{e,i}\hat{\alpha} + \beta_{s,i}(\hat{\alpha} - \mathbb{E}[\alpha]). \tag{51}$$

(In the case with symmetric products, this simplifies to $2\alpha r\beta_e + \alpha\beta_s(2r-1)$, so as we have known, increasing β_s benefits small firms if and only if $r < 1/2$.) When $\{\beta_{s,i}\}$ increase, the first term increases while the third one decreases as $\hat{\alpha} = \int r_i \alpha_i di < \mathbb{E}[\alpha] = \int \alpha_i di$. Consider two products i and k for which $\beta_{s,i}$ and $\beta_{s,k}$ increase by the same extent and $\alpha_i > \alpha_k$. Then it is clear that (51) is more likely to increase for product i. That is, the small firm that is better at generating data is more likely to suffer.

Consumer surplus is (up to a constant)

$$V = \int \left[\varphi_i(z_i) + \varepsilon\beta_{s,i}(\mathbb{E}[\alpha] - Q)\right] di \approx V_0 + \left[\int S_i\theta_i di + \mathbb{E}[\beta_s](\mathbb{E}[\alpha] - \hat{\alpha})\right]\varepsilon,$$

where $\varphi_i(z_i)$ is the surplus defined in the baseline case when there is no data cooperative, $V_0 = \int \varphi_i(r_i)di$ is consumer surplus when there are no data spillovers, and

$$S_i = \varphi_i'(r_i) = \frac{1 + r_i - \sigma_i(r_i)}{f_i} - \eta_{s,i}.$$

When $\{\beta_{s,i}\}$ increase in each market, the second term in the bracket must increase given $\mathbb{E}[\alpha] > \hat{\alpha}$, while the first term can decrease given S_i can be negative. This observation is also true when all products are symmetric, but there we have shown that the overall effect must be positive. However, with heterogeneous products, the opposite can be true. To see that, let us rewrite $(V - V_0)/\varepsilon$ as

$$\int \frac{S_i}{g_i}\left(\alpha_i(\hat{\beta}_e + \hat{\beta}_s) + (\beta_{e,i} + \beta_{s,i})\hat{\alpha} - \beta_{s,i}\mathbb{E}[\alpha]\right) di + \mathbb{E}[\beta_s](\mathbb{E}[\alpha] - \hat{\alpha})$$
$$= (\hat{\beta}_e + \hat{\beta}_s)\int \frac{S_i}{g_i}\alpha_i di + \hat{\alpha}\int \frac{S_i}{g_i}\beta_{e,i}di + (\mathbb{E}[\alpha] - \hat{\alpha})\int \beta_{s,i}\left(1 - \frac{S_i}{g_i}\right) di.$$

Suppose that $\beta_{s,i}$ increases by τ in each market. Then the impact on consumer surplus is τ times

$$\mathbb{E}[r]\int \frac{S_i}{g_i}\alpha_i di + (\mathbb{E}[\alpha] - \hat{\alpha})\int \left(1 - \frac{S_i}{g_i}\right) di.$$

This is negative if

$$\int \frac{S_i}{g_i}\alpha_i di < 0 \quad \text{and} \quad \int \frac{S_i}{g_i} di > 1.$$

This is possible, for example, when $\frac{S_i}{g_i} < 0$ and $\alpha_i > 0$ for $i \in [0, m]$, $\frac{S_i}{g_i} > 1$ and $\alpha_i = 0$ for $i \in (m, 1]$, and at the same time m is sufficiently small. This is the case when there are a small set of data-generating products with $S_i < 0$, and all the other are data-using products with $\frac{S_i}{g_i} > 1$.

Finally, we show that it is posible to have $S_i < 0$ and $\frac{S_i}{g_i} > 1$. Note that

$$S_i < 0 \iff \frac{1 + r_i - \sigma_i(r_i)}{f_i} < \eta_{s,i}.$$

In the linear Hotelling example, this condition holds if

$$\frac{\eta_{s,i}}{l_i} > 2 + \frac{3l_i - \eta_{s,i}}{3l_i - \bar{\eta}_i},$$

which is true if $\eta_{s,i}$ is sufficiently close to the upper bound $3l_i$ and is greater than $\eta_{e,i}$. On the other hand,

$$\frac{S_i}{g_i} > 1 \iff \eta_{e,i} f_i > 2 - r_i \left(1 + \frac{f_i'}{f_i^2}\right).$$

In the linear Hotelling example, this requires

$$\eta_{e,i} \geq 2l_i(2 - r_i) \iff \frac{3l_i - \eta_{s,i}}{3l_i - \bar{\eta}_i} \geq 4 - \frac{\eta_{e,i}}{l_i}.$$

This must hold when $\eta_{e,i}$ is above a threshold less than $3l_i$. To see that, notice that the left-hand side increases while the right-hand side decreases in $\eta_{e,i}$, and meanwhile, for any $\eta_{s,i} < 3l_i$, the above inequality holds at $\eta_{e,i} = 3l_i$. □

B Online Appendix

In this Online Appendix, we first generalize our baseline model to the case with a general form of cross-market data spillovers. We then report the case with two products to illustrate the additional complexity if we consider a discrete number of products instead of a continuum of them. Finally, we develop an alternative model where data helps reduce innovation cost rather than improve product quality directly.

B.1 A General Form of Cross-Market Data Spillovers

In this extension, we consider a more general form of cross-market data spillovers as suggested in modeling discussion (iv). When a consumer buys product i from the ecosystem, her surplus is

$$v_{e,i} - p_{e,i} + \int \beta_{ji} F_j(\Delta_j) dj + \epsilon_{e,i},$$

where β_{ji} measures how much the ecosystem's sales of product j can help improve its product i via the generated data, whereas if she buys from single-product firm i, her surplus is

$$v_{s,i} - p_{s,i} + \epsilon_{s,i}.$$

As before, here we define

$$\Delta_i = v_{e,i} - p_{e,i} + \int \beta_{ji} F_j(\Delta_j) dj - (v_{s,i} - p_{s,i}). \tag{52}$$

Our baseline setup is the special case with $\beta_{ji} = \alpha_j \beta_i$. To avoid trivial cases, we assume $\beta_{ij} > 0$ on a strictly positive measure of product pairs.

Let $\omega \geq 0$ denote the largest eigenvalue of the self-adjoint operator $(\beta + \beta^T)/2$, where $\beta = (\beta_{ji})$.[38] A useful observation is that, for any $\mathbf{t} = \{t_i\}_{i \in \mathcal{I}}$,

$$\int \int \beta_{ij} t_i t_j dj di = \int \int \beta_{ji} t_i t_j dj di = \int \int \frac{\beta_{ij} + \beta_{ji}}{2} t_i t_j dj di \leq \omega \int t_i^2 di, \tag{53}$$

where the first equality follows from exchanging the integral order.[39]

Consumer Problem Given firms' prices and investments, the equilibrium of the consumer choice game with rational expectations is characterized by $\{\Delta_i\}_{i \in \mathcal{I}}$ which solve (52) for each $i \in \mathcal{I}$.

Lemma 3. *For any given prices and quality investments, the consumer choice game has a unique equilibrium if*

$$\omega < \frac{1}{\max_i f_i(0)}. \tag{54}$$

[38] The operator $(\beta + \beta^T)/2$ is self-adjoint, and therefore all of its eigenvalues are real numbers. Also, $\omega \geq 0$ as each $\beta_{ji} \geq 0$. In addition, from (53), we have $\omega \leq \sup_{j,i} \beta_{ji}$, though this upper bound is often not tight.

[39] In the discrete case with n products, the inequality in (53) follows by observing that the n by n symmetric matrix $(\beta + \beta^T)/2 - \omega \mathbf{I}_n$ is negative semidefinte as its largest eigenvalue is zero. In the baseline setup with $\beta_{ji} = \alpha_j \beta_i$, we have shown in (40) that $2\omega = \chi = \mathbb{E}[\alpha\beta] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}$.

Proof. Note that $\{\Delta_i\}_{i\in\mathcal{I}}$ is a fixed point of a monotone mapping defined in (52) (recall that $\beta_{ij} \geq 0$). By Tarski's fixed-point theorem, an equilibrium of the consumer choice game must exist. We now prove uniqueness. Suppose in contrast that both $\{\Delta_i\}_{i\in\mathcal{I}}$ and $\{\Delta_i'\}_{i\in\mathcal{I}}$ are solutions to (52) and they differ in a strictly positive measure of i. Taking difference yields

$$\Delta_i - \Delta_i' = \int \beta_{ji}(F_j(\Delta_j) - F_j(\Delta_j'))dj.$$

Multiplying both sides by $(F_i(\Delta_i) - F_i(\Delta_i'))$ and integrating over i yields

$$\int (\Delta_i - \Delta_i')(F_i(\Delta_i) - F_i(\Delta_i'))di = \int\int \beta_{ji}(F_j(\Delta_j) - F_j(\Delta_j'))(F_i(\Delta_i) - F_i(\Delta_i'))djdi.$$

The right-hand side is at most $\omega \int (F_i(\Delta_i) - F_i(\Delta_i'))^2 di$ by (53), while the left-hand side, by the mean-value theorem, is

$$\int \frac{(F_i(\Delta_i) - F_i(\Delta_i'))}{f_i(\zeta_i)}(F_i(\Delta_i) - F_i(\Delta_i'))di \geq \frac{1}{\max_i f_i(0)} \int (F_i(\Delta_i) - F_i(\Delta_i'))^2 di$$

where ζ_i is a number between Δ_i and Δ_i'. Therefore,

$$\left(\frac{1}{\max_i f_i(0)} - \omega\right) \int (F_i(\Delta_i) - F_i(\Delta_i'))^2 di \leq 0.$$

Then condition (54) implies that $\int (F_i(\Delta_i) - F_i(\Delta_i'))^2 di = 0$, or $\Delta_i = \Delta_i'$ almost everywhere. This is a contradiction. □

Firm Problem We now turn to the firms' optimization problems. For the small firm i, its problem remains the same as in the baseline model, so its equilibrium price and quality choice take the same form as before: $p_{s,i} = (1 - F_i(\Delta_i))/f_i(\Delta_i)$ and $v_{s,i} = \eta_{s,i}(1 - F_i(\Delta_i))$.

The ecosystem's problem is more complicated. It aims to

$$\max_{\{p_{e,i},v_{e,i}\}_{i\in\mathcal{I}}} \int [p_{e,i}F_i(\Delta_i) - C_{e,i}(v_{e,i})]di$$

subject to a continuum of constraints in (52). We introduce a continuum of Lagrangian multipliers $\{\lambda_i\}$ and write

$$\mathcal{L} = \int \left[p_{e,i}F_i(\Delta_i) - \frac{v_{e,i}^2}{2\eta_{e,i}}\right] di + \int \lambda_i \left(v_{e,i} - p_{e,i} + \int \beta_{ji}F_j(\Delta_j)dj - (v_{s,i} - p_{s,i}) - \Delta_i\right) di.$$

Notice that

$$\int \lambda_i \int \beta_{ji} F_j(\Delta_j) dj di = \int \int \lambda_i \beta_{ji} di F_j(\Delta_j) dj = \int \underbrace{\left(\int \lambda_j \beta_{ij} dj \right)}_{\equiv \mu_i} F_i(\Delta_i) di$$

where the second equality is from exchanging the integral order, and $\mu_i = \int \lambda_j \beta_{ij} dj$ will be shown to be the subsidy to product i offered by the ecosystem. We can then rewrite $\mathcal{L}$ as

$$\mathcal{L} = \int \left[(p_{e,i} + \mu_i) F_i(\Delta_i) - \frac{v_{e,i}^2}{2\eta_{e,i}} + \lambda_i (v_{e,i} - p_{e,i} - (v_{s,i} - p_{s,i}) - \Delta_i) \right] di.$$

From the first-order conditions with respect to $(p_{e,i}, v_{e,i}, \Delta_i)$, we can derive

$$p_{e,i} = \frac{F_i(\Delta_i)}{f_i(\Delta_i)} - \mu_i, \quad v_{e,i} = \eta_{e,i} F_i(\Delta_i), \quad \lambda_i = F_i(\Delta_i).$$

Using the definition of μ_i, we then have

$$p_{e,i} = \frac{F_i(\Delta_i)}{f_i(\Delta_i)} - \int \beta_{ij} F_j(\Delta_j) dj.$$

When $\beta_{ij} = \alpha_i \beta_j$, this formula degenerates to what we had in the baseline model.

Define $G_i(\cdot)$ as (31) in the baseline model. Then the definition of Δ_i in (52), after substituting firms' prices and investments, can be written as

$$G_i(F_i(\Delta_i)) = -\eta_{s,i} + \int (\beta_{ij} + \beta_{ji}) F_j(\Delta_j) dj. \tag{55}$$

Now we need to directly deal with this system of a continuum of equations in $\{\Delta_i\}$. This is different from what we did in the baseline model where the problem boils down to a system of two equations in λ and Q. As before, once $\{\Delta_i\}$ is solved, we can pin down firms' equilibrium prices and quality investments.

Equilibrium Existence and Uniqueness We now extend Proposition 1 to this more general case.

Assumption 3. *For any $i \in \mathcal{I}$,*

$$\bar{\eta}_i + \omega < \frac{3}{2 f_i(0)}, \tag{C1}$$

and

$$\max\{\eta_{s,i}, \eta_{e,i} + \int (\beta_{ij} + \beta_{ji}) dj\} < \frac{3}{2f_i(0)}, \tag{C2}$$

and

$$\max\{\eta_{s,i}, \eta_{e,i} + 2\omega\} < \min_{-l_i \le x \le l_i} \frac{2 - \sigma_i(x)}{f_i(x)}. \tag{C3}$$

Remark. If we set $\beta_{ij} = \alpha_i \beta_j$ as in the baseline, this assumption reduces to Assumption 1. In particular, in (C1) and (C3), we just replace $\chi/2$ in (A1) and (A3) of Assumption 1 by ω. Also, one can check that condition (C3) implies (54) (which is required by equilibrium uniqueness of the consumer choice game).

Proposition 9. *Under Assumption 3, there exists a unique interior equilibrium.*

Proof. We first show that the system of equations in (55) has a unique solution under condition (C1). Given the continuum of equations, we cannot apply the Gale-Nikaido Theorem any more. Here we take a different approach. We rewrite the system (55) using ecosystem demands $\mathbf{z} = (z_i, i \in \mathcal{I})$: for each $i \in \mathcal{I}$,

$$G_i(z_i) = -\eta_{s,i} + \int (\beta_{ij} + \beta_{ji}) z_j dj. \tag{56}$$

Construct a function

$$\Gamma(\mathbf{z}) \equiv \int \left\{ \int_0^{z_i} (-G_i(t_i) - \eta_{s,i} + 2\omega t_i) dt_i \right\} di + \left[\int \int \beta_{ji} z_i z_j dj di - \omega \int z_i^2 di \right]. \tag{57}$$

This is strictly concave under the condition (C1). To see that, note that for each i, $\int_0^{z_i} (-G_i(t_i) - \eta_{s,i} + 2\omega t_i) dt_i$ is *strictly* concave in z_i as its second derivative is $-g_i(z_i) + 2\omega \le -(\frac{3}{f_i(0)} - 2\bar{\eta}_i) + 2\omega < 0$, where the first inequality used (34) and the second used (C1). Therefore, the first term of $\Gamma(\cdot)$ in (57) is strictly concave in $\mathbf{z}$. Meanwhile, the second term in (57) is a quadratic form of $\mathbf{z}$, and it is concave by the inequality (53). Hence, $\Gamma(\cdot)$, as the sum of these two terms, is strictly concave in $\mathbf{z}$.

Also, observe that

$$\begin{aligned} \frac{\partial \Gamma(\mathbf{z})}{\partial z_i} &= \{-G_i(z_i) - \eta_{s,i} + 2\omega z_i\} + \left[\int (\beta_{ij} + \beta_{ji}) z_j dj - 2\omega z_i \right] \\ &= -G_i(z_i) - \eta_{s,i} + \int (\beta_{ij} + \beta_{ji}) z_j dj. \end{aligned}$$

When $\mathbf{z}$ solves system (56), it must be a critical point of $\Gamma(\mathbf{z})$: for any $i \in \mathcal{I}$, $\frac{\partial \Gamma(\mathbf{z})}{\partial z_i} = 0$, which, by concavity of Γ, implies that $\mathbf{z}$ is in fact a global maximizer of $\Gamma(\cdot)$. However, a strictly concave function cannot have two distinct global maximizers, which implies the uniqueness of the solution to the system (56).[40]

Second, the proof that (C2) implies the solution $\{\Delta_i\}$ to (55) is interior is the same as in the baseline case, and is hence omitted.

Finally, we deal with the second-order conditions. The single-product firm's problem is exactly the same as before. For the ecosystem's problem, following a similar procedure as in the baseline, we work with quantities and rewrite its problem as

$$\max_{\mathbf{z}} \ \Pi_e(\mathbf{z}) = \int \left(\frac{\eta_{e,i}}{2} z_i - F_i^{-1}(z_i) - v_{s,i} + p_{s,i} \right) z_i di + \int \int \beta_{ji} z_i z_j dj di,$$

where we have substituted in the optimal investment decision $v_{e,i} = \eta_{e,i} z_i$. The proof for concavity under (C3) is the same as in the baseline analysis as long as we replace the inequality (40) there by the new inequality (53). $\square$

Effect of Data Regulation We can also extend Proposition 3 to this general setup. Consider a data policy that restricts the ecosystem's cross-product data usage (i.e, reducing β_{ij} in a positive measure of product pairs).

Proposition 10. *With general spillovers, data regulation that restricts cross-market data usage induces weakly lower Δ_i in all the markets and strictly lower Δ_i in a positive measure of markets.*

We can then immediately deduce that small firms weakly innovate more, set higher prices, sell more and earn more in all the markets and strictly so in a positive measure of markets. The opposite is true for the ecosystem's innovation and sales.

Proof. The proof exploits the monotonicity of the underlying system (55) and follows the lattice approach. Given β, we can define an operator $\mathcal{F}(\cdot;\beta)$ of $\mathbf{z} \in [0,1]^{\mathcal{I}}$:

$$\mathcal{F}(\mathbf{z};\beta) = (G_i^{-1}(-\eta_{s,i} + \int (\beta_{ij} + \beta_{ji}) z_j dj))_{i \in \mathcal{I}}$$

and rewrite system (56) as a fixed point equation of $\mathbf{z}$:

$$\mathcal{F}(\mathbf{z};\beta) = \mathbf{z}. \tag{58}$$

For a given β, the operator is monotone:

$$\mathbf{y}' \leq \mathbf{y}'' \implies \mathcal{F}(\mathbf{y}';\beta) \leq \mathcal{F}(\mathbf{y}'';\beta)$$

[40] This "potential function" approach can be also used to prove the uniqueness in the baseline model.

as $\beta_{ij} \geq 0$ for all i, j. As shown in the proof of Proposition 9, under (C1), the solution to (56) is unique, hence, this operator has a unique fixed point.

Now consider $\beta' = (\beta'_{ij})$ and $\beta = (\beta_{ij})$. Assume $\beta'_{ij} \leq \beta_{ij}$ for each $i, j \in \mathcal{I}$. Let $\mathbf{z}'$ and $\mathbf{z}$ denote the unique solution to system (58) under β' and β, respectively. We claim that

$$\mathbf{z}' \leq \mathbf{z}, \quad \text{i.e., } z'_i \leq z_i \text{ for all } i \tag{59}$$

By definition, $\mathbf{z} = \mathcal{F}(\mathbf{z}; \beta)$ and $\mathbf{z}' = \mathcal{F}(\mathbf{z}'; \beta')$. Since $\beta'_{ij} \leq \beta_{ij}$ and $z_i \geq 0$, we have, for any $\mathbf{y} \in [0,1]^{\mathcal{I}}$, $\mathcal{F}(\mathbf{y}; \beta) \geq \mathcal{F}(\mathbf{y}; \beta')$. Setting $\mathbf{y} = \mathbf{z}$ yields $\mathbf{z} = \mathcal{F}(\mathbf{z}; \beta) \geq \mathcal{F}(\mathbf{z}; \beta')$. Applying the operator $\mathcal{F}(\cdot; \beta')$ to both sides and exploiting its monotonicity, we obtain $\mathcal{F}(\mathbf{z}; \beta') \geq \mathcal{F}^2(\mathbf{z}; \beta')$. Keeping iterating yields

$$\mathbf{z} \geq \mathcal{F}(\mathbf{z}; \beta') \geq \mathcal{F}^2(\mathbf{z}; \beta') \geq \cdots \geq \mathcal{F}^n(\mathbf{z}; \beta') \geq \cdots$$

The monotone sequence $\{\mathcal{F}^n(\mathbf{z}; \beta')\}$ converges to a limit $\mathbf{z}^* \in [0,1]^{\mathcal{I}}$, and $\mathbf{z}^*$ must be $\mathbf{z}'$ as the operator $\mathcal{F}(\cdot; \beta')$ has a unique fixed point. Consequently, $\mathbf{z} \geq \mathbf{z}^* = \mathbf{z}'$.

Next we provide a strict version of the above result. Suppose, in addition to imposing $0 \leq \beta'_{ij} \leq \beta_{ij}$ for any i, j, we further assume strict inequalities, i.e, $0 \leq \beta'_{ij} < \beta_{ij}$, for a positive measure of product pairs. We claim that $z'_i < z_i$ for a positive measure of products. Using (56), we obtain, for each i,

$$G_i(z'_i) = -\eta_{s,i} + \int (\beta'_{ij} + \beta'_{ji}) z'_j dj \quad \text{and} \quad G_i(z_i) = -\eta_{s,i} + \int (\beta_{ij} + \beta_{ji}) z_j dj.$$

We have already shown that $z'_i \leq z_i$ for each i, which implies that

$$\int (\beta'_{ij} + \beta'_{ji}) z'_j dj \leq \int (\beta'_{ij} + \beta'_{ji}) z_j dj.$$

Therefore,

$$G_i(z_i) - G_i(z'_i) \geq \int (\beta_{ij} + \beta_{ji}) z_j dj - \int (\beta'_{ij} + \beta'_{ji}) z_j dj \tag{60}$$

Multiplying by z_i on both sides and integrating over i yield

$$\int (G_i(z_i) - G_i(z'_i)) z_i di \geq \int \int \underbrace{\left((\beta_{ij} + \beta_{ji}) - (\beta'_{ij} + \beta'_{ji})\right)}_{\geq 0} z_i z_j di dj > 0$$

The last strict inequality is because $\beta_{ij} - \beta'_{ij} > 0$ on a positive measure of product pairs, and $z_i > 0$ for each i in an interior equilibrium. From $\int (G_i(z_i) - G_i(z'_i)) z_i di > 0$, we obtain that, on a positive measure of products i, $G_i(z_i) - G_i(z'_i)) > 0$, or equivalently $z_i > z'_i$. $\square$

B.2 The Two-Product Case

This section aims to demonstrate that the case of a finite number of products is more complicated to deal with than our continuum case. Suppose that the ecosystem supplies only two products $i = 1, 2$, and in each product market there is also a different single-product competitor. Let F_i be the preference distribution and $(\eta_{e,i}, \eta_{s,i})$ be the innovation cost parameters in market i.

Let Q_i denote the ecosystem's sales in market i. Define

$$\Delta_i = v_{e,i} - p_{e,i} + \beta_{ii}Q_i + \beta_{ji}Q_j - (v_{s,i} - p_{s,i} + \hat{\beta}_{ii}(1 - Q_i)),$$

where β_{ii} and $\hat{\beta}_{ii}$ capture respectively the within-market data effects for the ecosystem and the small firm, and β_{ji} captures the cross-market data effect for the ecosystem. (In our continuum framework, the amount of within-market data was negligible compared to the amount of cross-market data, so the within-market data effects were ignored.) If we separate data generation from data usage as in the baseline model, we have $\beta_{ii} = \alpha_i\beta_{e,i}$, $\beta_{ji} = \alpha_j\beta_{e,i}$, and $\hat{\beta}_{ii} = \alpha_i\beta_{s,i}$. Assuming rational expectations, the consumer choice game solves

$$Q_1 = F_1(\Delta_1) \quad \text{and} \quad Q_2 = F_2(\Delta_2).$$

The ecosystem's Lagrange problem is

$$\max \sum_{i=1}^{2} \left[p_{e,i}F_i(\Delta_i) - \frac{v_{e,i}^2}{2\eta_{e,i}} + \lambda_i(F_i(\Delta_i) - Q_i) \right].$$

From the first-order conditions with respect to $p_{e,i}$ and $v_{e,i}$, we derive

$$F_i - p_{e,i}f_i - \lambda_i f_i = 0 \implies p_{e,i} = \frac{F_i}{f_i} - \lambda_i,$$

and

$$p_{e,i}f_i - \frac{v_{e,i}}{\eta_{e,i}} + \lambda_i f_i = 0 \implies v_{e,i} = \eta_{e,i}F_i.$$

(We have suppressed the dependent variable Δ_i in both f_i and F_i.) The first-order condition with respect to Q_i is

$$p_{e,i}f_i(\beta_{ii} + \hat{\beta}_{ii}) + p_{e,j}f_j\beta_{ij} + \lambda_i(f_i(\beta_{ii} + \hat{\beta}_{ii}) - 1) + \lambda_j f_j\beta_{ij} = 0,$$

from which we derive

$$\lambda_i = (\beta_{ii} + \hat{\beta}_{ii})F_i + \beta_{ij}F_j$$

by using the first-order conditions of price. Therefore, the ecosystem's equilibrium price and investment choices take a qualitatively similar form as in the continuum framework. The only difference is that now the within-market data effect is an additional force for the ecosystem to subsidize data collection.

Each single-product firm's problem, however, is more complicated since their price and quality affect both Q_1 and Q_2 and so they face a constrained optimization problem similar as the ecosystem. Let us consider single-product firm 1. Its Lagrange problem is

$$\max \ p_{s,1}[1 - F_1(\Delta_1)] - \frac{v_{s,1}^2}{2\eta_{s,1}} + \sum_{i=1}^{2} \mu_i (Q_i - F_i(\Delta_i)),$$

where (μ_1, μ_2) are the Lagrangian multipliers of firm 1. From the first-order conditions with respect to $p_{s,i}$ and $v_{s,i}$, we derive

$$p_{s,1} = \frac{1 - F_1}{f_1} - \mu_1 \quad \text{and} \quad v_{s,1} = \eta_{s,1}(1 - F_1).$$

Due to the within-market data effect, now a single-product firm will also subsidize its data collection, which is captured by μ_1. From the first-order conditions with respect to Q_1 and Q_2, we can derive

$$\begin{aligned} \mu_1 &= (1 - F_1) \left(\frac{\beta_{12}\beta_{21} f_2}{1 - (\beta_{22} + \hat{\beta}_{22}) f_2} + \beta_{11} + \hat{\beta}_{11} \right), \\ \mu_2 &= \frac{1 - F_1}{1 - (\beta_{22} + \hat{\beta}_{22}) f_2} \beta_{21}. \end{aligned}$$

(We first derived μ_2 from the first-order condition with respect to Q_2 by using the pricing first-order conditions; using μ_2 we then derive μ_1 from the first-order condition with respect to Q_2.) Notice that the small firms' pricing formula is significantly more complicated than in the continuum case. The last two terms $\beta_{11} + \hat{\beta}_{11}$ in μ_1 are due to the within-market data effect: lowering $p_{s,1}$ helps small firm 1 acquire data in market 1, which not only improves small firm 1's own quality but also reduces the ecosystem's quality. The first term in μ_1 captures the cross-market data effect: lowering $p_{s,1}$ reduces the ecosystem's data from market 1, which weakens its position in market 2 and also reduces its data there as well, which in turns helps small firm 1 in market 1.

The Δ_i equation now becomes

$$\Delta_i = (\eta_{e,i} + 2\beta_{ii} + \hat{\beta}_{ii})F_i + \frac{1-2F_i}{f_i} + (\beta_{ij} + \beta_{ji})F_j - (1-F_i)\left(\eta_{s,i} + \beta_{ii} + 2\hat{\beta}_{ii} + \frac{\beta_{ij}\beta_{ji}f_j}{1-(\beta_{jj}+\hat{\beta}_{jj})f_j}\right),$$

which is also much more involved than in the continuum case.

By using the quantity approach, we can derive a qualitatively similar sufficient condition as in the continuum case for the ecosystem's problem to be concave. Nevertheless, it becomes much more challenging in finding a clean sufficient condition for a single-product firm's problem to be concave.

B.3 Data and Innovation Costs

In our baseline analysis, data can be used to directly improve product quality. Here we consider a different scenario where data reduces a firm's innovation cost, and so can indirectly improve product quality. For simplicity we focus on the baseline model (i.e., the case without data sharing or a data cooperative).

Consider the following set-up. In each market $i \in \mathcal{I}$ the ecosystem and relevant single-product firm again compete in price and quality/innovation. However now the fixed costs of investment for product i are respectively

$$C_{e,i}(v_{e,i}) = \frac{v_{e,i}^2}{2(\eta_{e,i}+\beta_i Q)} \quad \text{and} \quad C_{s,i}(v_{s,i}) = \frac{v_{s,i}^2}{2\eta_{s,i}}$$

where Q again denotes the total amount of data and is given by

$$Q = \int \alpha_i z_i di,$$

with z_i again denoting ecosystem sales in market i. If a consumer buys from the ecosystem in market i her surplus is $v_{e,i} - p_{e,i} + \epsilon_{e,i}$, while if she buys from the single-product firm i her surplus is the same as before, that is, $v_{s,i} - p_{s,i} + \epsilon_{s,i}$. All other aspects of the model remain the same as in the baseline.

Equilibrium Analysis We again start by solving a consumer's problem. Define

$$\Delta_i = (v_{e,i} - p_{e,i}) - (v_{s,i} - p_{s,i}), \tag{61}$$

and note that in market i the ecosystem sells to $F_i(\Delta_i)$ consumers. Unlike in the main analysis, now Δ_i does not (directly) depend on Q, and so the consumer choice game has a unique equilibrium without the need for any conditions on model parameters.

Now turn to the firms' optimization problems. Single-product firm i's problem takes the same form as in the baseline analysis:

$$\max_{p_{s,i}, v_{s,i}} \quad p_{s,i}[1 - F_i(\Delta_i)] - \frac{v_{s,i}^2}{2\eta_{s,i}}.$$

Hence single-product firm i's optimal price and quality investment also take the same form as in the baseline analysis:

$$p_{s,i} = \frac{1 - F_i(\Delta_i)}{f_i(\Delta_i)} \quad \text{and} \quad v_{s,i} = \eta_{s,i}[1 - F_i(\Delta_i)]. \tag{62}$$

The ecosystem's problem is different, however, since Q affects its investment costs rather than the surpluses it offers consumers. Following the same procedure as in the baseline analysis, the ecosystem's Lagrangian is:

$$\mathcal{L} = \int \left[p_{e,i} F_i(\Delta_i) - \frac{v_{e,i}^2}{2(\eta_{e,i} + \beta_i Q)} \right] di + \lambda \left[\int \alpha_i F_i(\Delta_i) di - Q \right].$$

Taking first-order conditions with respect to $(p_{e,i}, v_{e,i})$ and solving them, we obtain that on product i:

$$p_{e,i} = \frac{F_i(\Delta_i)}{f_i(\Delta_i)} - \lambda \alpha_i \quad \text{and} \quad v_{e,i} = (\eta_{e,i} + \beta_i Q) F_i(\Delta_i), \tag{63}$$

which are qualitatively the same as in the benchmark analysis. Taking a first-order condition with respect to Q and using the above expression for $v_{e,i}$ to simplify it gives

$$\lambda = \int \frac{\beta_i F_i(\Delta_i)^2}{2} di.$$

To understand this expression, notice that starting from the optimal $\{p_{e,i}, v_{e,i}\}_{i \in \mathcal{I}}$, the above captures how much the ecosystem saves on investment costs if given a bit more data. As in the baseline model, we have now solved for all prices and investments as a function of the $\{\Delta_i\}_{i \in \mathcal{I}}$. This allows us to rewrite equation (61) as

$$\Delta_i = 2\overline{\eta}_i F_i(\Delta_i) + \frac{1 - 2F_i(\Delta_i)}{f_i(\Delta_i)} - \eta_{s,i} + \alpha_i \lambda + \beta_i Q F_i(\Delta_i), \tag{64}$$

where recall that

$$Q = \int \alpha_i F_i(\Delta_i) di \quad \text{and} \quad \lambda = \int \frac{\beta_i F_i(\Delta_i)^2}{2} di. \tag{65}$$

There are two differences with the baseline analysis. First, the value of data λ takes a different form, which is now quadratic rather than linear in $F_i(\Delta_i)$. Second, the Δ_i equation is different than the corresponding expression in the baseline analysis (i.e., equation (10)) because the final term on the righthand side is now $\beta_i Q F_i(\Delta_i)$ whereas in the baseline model it is $\beta_i Q$. Both differences arise because in this extension data improves product quality only indirectly (and also non-linearly) by reducing the innovation cost, and because ecosystem investment on a given product is proportional to its sales of that product.

We now turn to existence and uniqueness of equilibrium. For convenience, define

$$\xi \equiv \underbrace{\mathbb{E}[\alpha\beta] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}}_{\equiv \chi} + \max_{i \in \mathcal{I}} \beta_i \mathbb{E}[\alpha].$$

In the subsequent analysis, we make the following assumption on primitives:

Assumption 4. *For any $i \in \mathcal{I}$,*

$$\xi + 2\bar{\eta}_i < \frac{3}{f_i(0)} \tag{D1}$$

and

$$\max\left\{\eta_{s,i}, \eta_{e,i} + \alpha_i \frac{\mathbb{E}[\beta]}{2} + \beta_i \mathbb{E}[\alpha]\right\} < \frac{3}{2f_i(0)} \tag{D2}$$

and

$$\max\{\eta_{s,i}, \eta_{e,i} + \xi\} < \min_{-l_i \le x \le l_i} \frac{2 - \sigma_i(x)}{f_i(x)}. \tag{D3}$$

This is qualitatively like Assumption 1 in the baseline model, and thus also holds provided product differentiation is sufficiently large, i.e., provided in each market i we have $f_i(0)$ sufficiently small. We can then state the following result:

Proposition 11. *Under Assumption 4, there exists a unique interior equilibrium, and it is characterized by* (62), (63), (64) *and* (65).

Proof. We first prove that the putative equilibrium is interior. Notice that, using equation (64), $\Delta_i > -l_i$ requires that $-l_i < \frac{1}{f_i(-l_i)} - \eta_{s,i} + \alpha_i \lambda$, while $\Delta_i < l_i$ requires that

$l_i > \eta_{e,i} - \frac{1}{f_i(l_i)} + \alpha_i\lambda + \beta_i Q$. Using the same steps as in the proof of Proposition 1 from the baseline analysis, as well as the fact that $0 \le \lambda \le \frac{\mathbb{E}[\beta]}{2}$ and $0 \le Q \le \mathbb{E}[\alpha]$, these conditions are satisfied provided (D2) holds.

Next, consider the second-order conditions. Since single-product firm i's problem has the same form as in the baseline analysis, exactly the same steps can be used to show that it does not have a profitable unilateral deviation provided

$$\eta_{s,i} < \min_{-l_i \le x \le l_i} \frac{2 - \sigma_i(x)}{f_i(x)}.$$

Note that this holds given condition (D3). For the ecosystem, we begin by following the same approach as in the baseline analysis and rewrite its profit as a function of sales and qualities:

$$\Pi_e(\{z_i, v_{e,i}\}_{i\in\mathcal{I}}) \equiv \int \left[(v_{e,i} - F_i^{-1}(z_i) - v_{s,i} + p_{s,i}) z_i - \frac{v_{e,i}^2}{2[\eta_{e,i} + \beta_i(\int \alpha_i z_i di)]} \right] di,$$

where we have used the definition of Q. Substituting $v_{e,i} = z_i[\eta_{e,i} + \beta_i(\int \alpha_i z_i di)]$ into the objective function, the problem becomes

$$\max_{\mathbf{z}} \Pi_e(\mathbf{z}) = \int \left[\frac{1}{2} \left[\eta_{e,i} + \beta_i \left(\int \alpha_i z_i di \right) \right] z_i - F_i^{-1}(z_i) - v_{s,i} + p_{s,i} \right] z_i di.$$

We claim that $\Pi_e(\mathbf{z})$ is concave in $\mathbf{z} \in [0,1]^{\mathcal{I}}$ if the following holds for each $i \in \mathcal{I}$:

$$\eta_{e,i} + \xi < \min_{z_i \in [0,1]} [z_i F_i^{-1}(z_i)]'' = \min_{x \in [-l_i, l_i]} \frac{2 - \sigma_i(x)}{f_i(x)}. \tag{66}$$

Note that this holds given condition (D3). In the baseline model $\Pi_e(\mathbf{z})$ contains quadratic terms in z_i, whereas here it contains cubic terms in z_i, so the proof now takes a different approach compared to the proof of second-order conditions in the baseline model. In particular, to prove the above claim, we decompose $\Pi_e(\mathbf{z})$ into the sum of two terms:

$$\Pi_e(\mathbf{z}) = \int \left[\frac{\eta_{e,i} + \xi}{2} z_i - F_i^{-1}(z_i) - v_{s,i} + p_{s,i} \right] z_i di + \theta(\mathbf{z}),$$

where the first term, as in the baseline model, is additive across markets and is concave under the condition in (66), and the second term

$$\theta(\mathbf{z}) \equiv \overbrace{\int \alpha_i z_i di}^{=Q} \times \overbrace{\int \frac{1}{2} \beta_i z_i^2 di}^{=\lambda} - \xi \int \frac{1}{2} z_i^2 d_i. \tag{67}$$

is concave. To show the latter, we prove that its Hessian matrix, $\nabla^2\theta(\mathbf{z})$ is negative definite at any $\mathbf{z} \in [0,1]^{\mathcal{I}}$, in the sense that, for any $\mathbf{x}$, the quadratic form

$$\mathbf{x}'\nabla^2\theta(\mathbf{z})\mathbf{x} = \int\int \frac{\partial^2\theta(\mathbf{z})}{\partial z_i \partial z_j} x_i x_j di dj \leq 0.$$

Direct computation yields

$$\frac{\partial\theta(\mathbf{z})}{\partial z_i} = \alpha_i\lambda + \beta_i z_i Q - \xi z_i$$

and

$$\frac{\partial^2\theta(\mathbf{z})}{\partial z_i \partial z_j} = \alpha_i(\beta_j z_j) + (\beta_i z_i)\alpha_j + \beta_i Q\delta_{i,j} - \xi\delta_{i,j}$$

where $\lambda = \mathbb{E}[\frac{1}{2}\beta z^2] = \int \frac{1}{2}\beta_i z_i^2 di$ and $Q = \mathbb{E}[\alpha z] = \int \alpha_i z_i di$, and $\delta_{i,j}$ is the Kronecker delta function. Therefore,

$$\begin{aligned}
\mathbf{x}'\nabla^2\theta(\mathbf{z})\mathbf{x} &= \int\int \frac{\partial^2\theta(\mathbf{z})}{\partial z_i \partial z_j} x_i x_j di dj \\
&= \int\int (\alpha_i(\beta_j z_j) + (\beta_i z_i)\alpha_j) x_i x_j di dj + Q\int \beta_i x_i^2 di - \xi\int x_i^2 di \\
&\leq \int\int (\alpha_i\beta_j + \alpha_j\beta_i)|x_i||x_j| di dj + \left(\max_{i\in\mathcal{I}} \beta_i\right)\mathbb{E}[\alpha]\int x_i^2 di - \xi\int x_i^2 di \\
&= 2\int\int \alpha_i\beta_j |x_i||x_j| di dj + \left(\max_{i\in\mathcal{I}} \beta_i\right)\mathbb{E}[\alpha]\int x_i^2 di - \xi\int x_i^2 di \\
&\leq 2 \times \frac{\mathbb{E}[\alpha\beta] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]}}{2}\int |x_i|^2 di + \left(\max_{i\in\mathcal{I}} \beta_i\right)\mathbb{E}[\alpha]\int x_i^2 di - \xi\int x_i^2 di \\
&= \left[\mathbb{E}[\alpha\beta] + \sqrt{\mathbb{E}[\alpha^2]\mathbb{E}[\beta^2]} + \left(\max_{i\in\mathcal{I}} \beta_i\right)\mathbb{E}[\alpha] - \xi\right]\int x_i^2 di = 0,
\end{aligned}$$

where the first inequality uses $z_i \leq 1, \alpha_i \geq 0, \beta_i \geq 0$, and $Q \leq \mathbb{E}[\alpha]$, the second inequality uses the χ-inequality (40) from the existence proof in the baseline analysis, and the last equality follows from the definition of ξ.

Finally, we establish uniqueness of equilibrium. Notice that, using the function $G_i(x)$ defined in equation (31), as well as the fact that $z_i = F_i(z_i)$, we can rewrite equation (64) as

$$G_i(z_i) = -\eta_{s,i} + \alpha_i\lambda + \beta_i Q z_i, \quad \forall i \in \mathcal{I} \tag{68}$$

and rewrite equation (65) as

$$Q = \int \alpha_i z_i di \quad \text{and} \quad \lambda = \int \frac{\beta_i}{2} z_i^2 di.$$

Notice that for given (λ, Q) equation (68) has a unique (stable) solution z_i if $G_i'(z_i) > \beta_i Q$, which is implied by condition (D1) because $G_i' \geq \frac{3}{f_i(0)} - 2\bar{\eta}_i$ and because $\beta_i Q \leq (\max_{i \in \mathcal{I}} \beta_i)\mathbb{E}[\alpha]$. We now prove that the solution $\mathbf{z}$ is unique. Since it is no longer convenient to use the Gale-Nikaido Theorem here, we develop a different approach. A key observation is that system (68) can be viewed as a critical point of a function, which can be verified to be strictly concave. In particular, define the following function:

$$\Gamma(\mathbf{z}) \equiv -\int \left\{ \int_0^{z_i} (G_i(t_i) - \xi t_i) dt_i \right\} di - \left[\int \eta_{s,i} z_i di \right] + \theta(\mathbf{z}),$$

where the third term θ is defined in (67). For each i, the term $\int_0^{z_i} (G_i(t_i) - \xi t_i) dt_i$ is *strictly* convex in z_i as its second derivative is $G_i'(z_i) - \xi \geq \frac{3}{f_i(0)} - 2\bar{\eta}_i - \xi > 0$ by condition (D1). Therefore, the first term in $\Gamma(\mathbf{z})$ is strictly concave in $\mathbf{z}$. Meanwhile, the second term is linear in $\mathbf{z}$ and the third term $\theta(\mathbf{z})$, as we have shown before, is concave. Hence $\Gamma(\mathbf{z})$ is strictly concave in $\mathbf{z} \in [0,1]^{\mathcal{I}}$. Next, observe that

$$\alpha_i \lambda + \beta_i z_i Q = \frac{\partial Q}{\partial z_i} \lambda + \frac{\partial \lambda}{\partial z_i} Q = \frac{\partial (Q\lambda)}{\partial z_i}.$$

Hence we have

$$\frac{\partial \Gamma(\mathbf{z})}{\partial z_i} = -(G_i(z_i) - \xi z_i) - \eta_{s,i} + \frac{\partial \theta(\mathbf{z})}{\partial z_i} = -(G_i(z_i) + \eta_{s,i}) + (\alpha_i \lambda + \beta_i z_i Q).$$

When $\mathbf{z}$ solves system (68), it must be a critical point of $\Gamma(\cdot)$: for any $i \in \mathcal{I}$, $\frac{\partial \Gamma(\mathbf{z})}{\partial z_i} = 0$, which, by concavity of Γ, implies that $\mathbf{z}$ is in fact a global maximizer of $\Gamma(\cdot)$. However, a strictly concave function cannot have two distinct global maximizers, which implies the uniqueness of the solution to the system (68). $\square$

Cross-Market Comparison As in the baseline analysis, consider two markets which are identical except for their data parameters α_i, β_i. Then, from equation (64), in market i we have

$$\Delta_i = 2\bar{\eta} F(\Delta_i) + \frac{1 - 2F(\Delta_i)}{f(\Delta_i)} - \eta_s + \alpha_i \lambda + \beta_i Q F(\Delta_i).$$

Assumption 4 ensures that the left-hand side of this equation increases faster in Δ_i than does the right-hand side. Since $\lambda > 0$ and $Q > 0$, this implies that, as in the baseline model, the ecosystem sells more (other things equal) in markets with higher α_i and β_i. It then follows from equations (62) and (63) that in markets with higher α_i

and β_i, the corresponding small firm charges less and invests less, while the ecosystem invests more. Turning to the ecosystem's price, it is clear from (63) that, other things equal, the ecosystem charges more in markets with higher β_i, but that it is ambiguous whether it charges more or less in markets with higher α_i. To explore this further, one can check that in the linear Hotelling case

$$p_{e,i} = \frac{2l(3l - \eta_s + \alpha_i \lambda)}{6l - \eta_e - \eta_s - \beta_i Q} - \alpha_i \lambda.$$

Notice that $6l - \eta_e - \eta_s - \beta_i Q > 6l - \eta_e - \eta_s - \beta_i \mathbb{E}\alpha > 0$ where the first inequality uses (65) and the second inequality uses (D2). One can then check that $p_{e,i}$ *increases* in α_i if and only if $2\bar{\eta} > 4l - \beta_i Q$, which is compatible with Assumption 4, and which has a similar flavor to the corresponding condition from the benchmark analysis.

Data Regulation Consider again regulation that reduces α_i or β_i in a positive measure of markets. Consistent with Proposition 3 from the main analysis, we find that:

Proposition 12. *(i) Data regulation induces the ecosystem to (weakly) sell less and innovate less in all markets, and induces all single-product firms to (weakly) innovate more, set higher prices, sell more, and earn more profit.*
(ii) It benefits consumers if

$$\eta_{s,i} > (1 + F_i(\Delta_i) - \sigma_i(\Delta_i))/f_i(\Delta_i)$$

in each market $i \in \mathcal{I}$, and harms consumers if the opposite is true in each market $i \in \mathcal{I}$.

Proof. First, we rewrite (68) equivalently as a fixed-point of an operator that maps any $\mathbf{z} = \{z_j, j \in \mathcal{I})\}$ in $[0,1]^{\mathcal{I}}$ to

$$\left\{ G_j^{-1}\left(-\eta_{s,j} + \alpha_j \left(\int \frac{\beta_i}{2} z_i^2 di \right) + \beta_j z_j \left(\int \alpha_i z_i di \right) \right),\ j \in \mathcal{I} \right\} \tag{69}$$

in $[0,1]^{\mathcal{I}}$. Given $\alpha_i \geq 0, \beta_i \geq 0, z_i \geq 0$ and monotonicity of $G_i(\cdot)$, this operator is monotone in $\mathbf{z}$. Furthermore, for a fixed $\mathbf{z}$, the operator is monotone in the data parameters $\{\alpha_i, \beta_i, i \in \mathcal{I}\}$. By following the same argument as in the proof of Proposition 10 we can show that data regulation in this extension always reduces z_i (at least weakly) for each i.[41] It then follows that λ and Q also decrease with the regulation.

[41] In fact, ecosystem sales in market i strictly decrease if either $\alpha_i + \beta_i > 0$, or if $\alpha_i + \beta_i$ strictly goes down after the regulation.

The ecosystem's innovation (from equation (63)) is $v_{e,i} = (\eta_{e,i} + \beta_i Q)F_i(\Delta_i)$. Since data regulation weakly reduces $\beta_i \geq 0$, $Q \geq 0$ and Δ_i in each market, it also reduces ecosystem innovation in each market.

Now consider the impact on single-product firms. Note that $p_{s,i}$ and $v_{s,i}$ in equation (62) both decrease in Δ_i, and hence increase (weakly) due to the regulation. Meanwhile, it is easy to see that a single-product firm's profit takes the same form as in the baseline model (i.e., is the same as in equation (17)). Following the same logic as in the baseline model, condition (D3) therefore implies that single-product firm i's profit decreases in Δ_i, and thus (weakly) increases due to the regulation.

Finally, consider consumer surplus. Notice that, up to a constant, consumer surplus in market i takes the usual form:

$$V_i = [1 - F_i(\Delta_i)] \left[\eta_{s,i} - \frac{1}{f_i(\Delta_i)} \right] + \int_0^{\Delta_i} F_i(x)dx$$

which decreases in Δ_i if and only if the condition in the proposition holds. □

As in the benchmark analysis, data regulation reduces the value λ and volume Q of data collected by the ecosystem. Regulation also induces the ecosystem to sell (weakly) less in each market, which in turn explains the impacts on ecosystem investment and on small firms' investments, prices, and profits described in Proposition 12. Moreover, as in the benchmark analysis, data regulation has an ambiguous impact on ecosystem profit and consumer surplus;[42] as before, regulation is more likely to benefit consumers when the $\eta_{s,i}$ are relatively large, i.e., when small firms are relatively efficient at investing.

To explore the last point in more detail, Figures 5 and 6 depict ecosystem profit and consumer surplus in the linear Hotelling case with symmetric products.[43] The

[42]The ecosystem's profit in market i can be written as

$$\Pi_{e,i} = F_i(\Delta_i) \left[\frac{F_i(\Delta_i)}{f_i(\Delta_i)} - \alpha_i \lambda \right] - \frac{\eta_{e,i} + \beta_i Q}{2} F_i(\Delta_i)^2$$

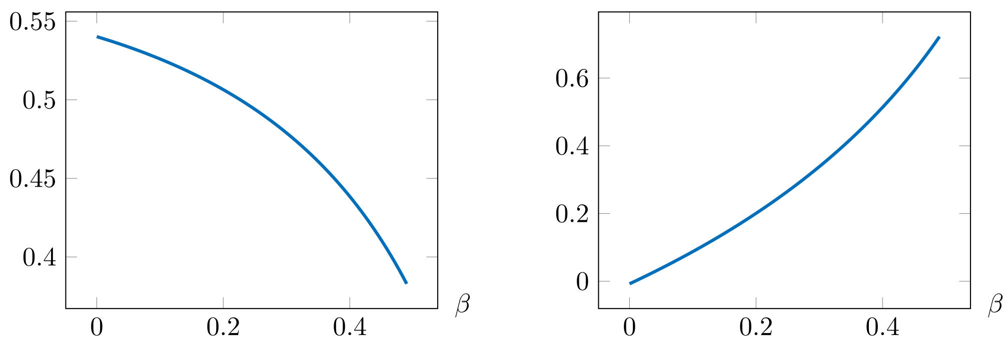
Aggregating over all markets in $\mathcal{I}$ and using the expressions for λ and Q in (65) this simplifies to

$$\Pi_e = \int F_i(\Delta_i)^2 \left[\frac{1}{f_i(\Delta_i)} - \frac{\eta_{e,i}}{2} \right] - 2\lambda Q.$$

which looks similar to profit in the baseline model (in equation (18)) except for the 2 in front of the subsidy term and the fact that λ takes a different form. As in the baseline model, regulation has an ambiguous effect on ecosystem profit due to this subsidy term.

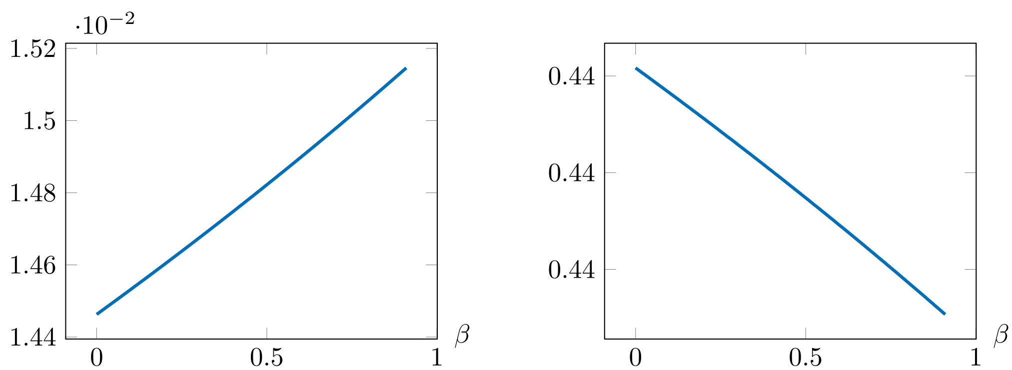
[43]Note that in the linear Hotelling model with symmetric products $\sigma(x) = 0$ and $\xi = 3\beta$. Hence (D1)

effect of regulation is qualitatively the same as in the corresponding Figures 2 and 3 from the baseline analysis. (Indeed, each pair of figures uses the same underlying model parameters.) Specifically, in the first plot, small firms are relatively inefficient at investing: regulation that reduces β only induces a relatively small increase in small-firm innovation, and so it benefits the ecosystem but harms consumers. In the second plot, however, small firms are relatively efficient at investing: regulation that reduces β induces a relatively large increase in small-firm innovation, and so it benefits consumers but harms the ecosystem.



(a) Ecosystem Profit (b) Consumer Surplus

Figure 5: The impact of β when data affects the ecosystem's innovation cost
(products are symmetric, F is uniform on $[-1, 1]$, and $\alpha = 1, \eta_e = 2.25, \eta_s = 0.25$)



(a) Ecosystem Profit (b) Consumer Surplus

Figure 6: The impact of β when data affects the ecosystem's innovation cost
(products are symmetric, F is uniform on $[-1, 1]$, and $\alpha = 1, \eta_e = 0.5, \eta_s = 2.75$)

becomes $2\bar{\eta} + 3\beta < 3/f(0)$, (D2) becomes $\max\{\eta_s, \eta_e + 3\beta/2\} < 3/[2f(0)]$, and (D3) becomes $\max\{\eta_s, \eta_e + 3\beta\} < 2/f(0)$.