WORKING PAPERS

N° 1621

February 2026

"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 2026

Abstract

This paper develops a framework in which a multiproduct ecosystem competes with multiple single-product firms in both price and innovation. The ecosystem can use data from one product to improve the quality of its other products. We use the framework to study three regulatory policies aimed at leveling the playing field. Restricting the ecosystem's cross-product data usage, or forcing it to share data with single-product firms, benefits those firms and induces them to innovate more. However, these policies also dampen the ecosystem's incentive to collect data and innovate, potentially raising prices. Consumers are better off only when single-product firms are sufficiently good at innovating. Facilitating data exchange between single-product firms via a data cooperative can backfire and harm them, because it induces the ecosystem to price more aggressively. For both the data-sharing and data-cooperative policies, there exist data-compensation schemes such that consumers are better off compared to no regulation.

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

JEL classification: D43, L13, L51

*This paper is a substantially revised version of an earlier draft circulated under the same title in 2025. We are grateful to Nageeb Ali, Mark Armstrong, Yeon-Koo Che, Jay Pil Choi, Rahul Deb, Kai Fischer, Sanjeev Goyal, Michael Grubb, Paul Heidhues, Meg Meyer, Aniko Oery, David Ronayne, Tom Ross, Ali Shourideh, Guofu Tan, Tat-How Teh, Alex Teytelboym, Julian Wright, Yaron Yehezkel, 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 the AI Interdisciplinary Institute ANITI (grant ANR-23-IACL-0002), 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

Big tech companies like Google, Apple, Meta, Amazon, and Microsoft operate as "digital ecosystems," offering a wide range of products and services that are interconnected at the technological or data level. As illustrated in Figure 1, these ecosystems typically compete not just with each other, but also with firms that specialize in particular products or services. Moreover, many of these specialized rivals, such as Spotify, PayPal, eBay, and OpenAI, are significant competitors that play important roles in their respective markets.

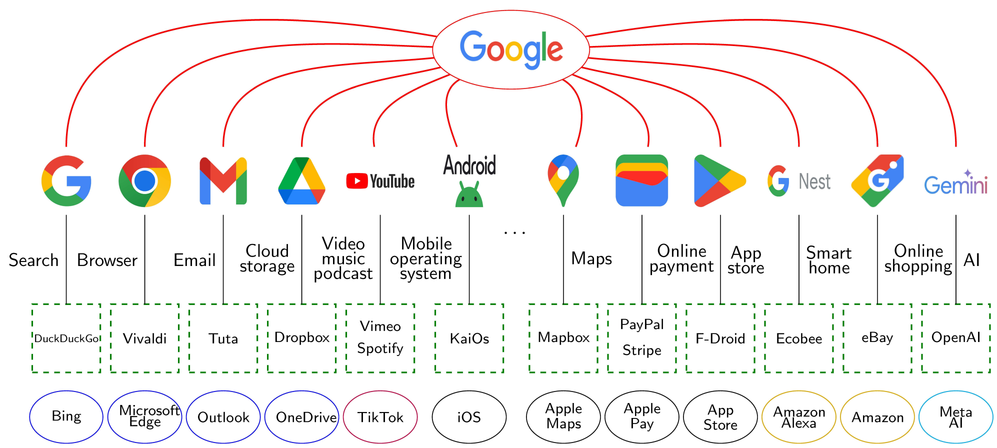


Figure 1: Google's ecosystem and some of its competitors
(specialized competitors in boxes; competing products from other ecosystems in ovals)

There are widespread concerns that these big tech companies have become too big and too powerful, putting competitors at a disadvantage and hurting both smaller businesses and consumers.[1] There are many reasons why these companies may have a competitive advantage over smaller firms, including brand recognition, better access to capital, and superior infrastructure. However, one particular advantage stems from the vast amount of data they collect, as well as their ability to use data gleaned in one product market to improve their offering in other markets.[2] For example, Google can

[1]See https://shorturl.at/PQVs2 for a U.S. survey. Similar concerns have been raised about big tech companies like Tencent, Alibaba, and ByteDance in China. See https://shorturl.at/Z7WRu.

[2]For instance, 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.

use location data from Google Maps, trends data from Google Shopping, and viewing data from Youtube, to enhance the relevance of its search and advertising results, in a way that a specialized competitor like DuckDuckGo cannot. Similarly, it can use data from sources like Google search and YouTube to train Gemini, enabling it to catch up with and even leapfrog its specialized competitor, OpenAI.[3] Meanwhile, ecosystems in China 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.

Specialized firms' inability to use cross-market data may make it harder for them to compete with big tech companies—dampening their incentives to innovate and expand, and ultimately harming final consumers (see, e.g., ACCC, 2023). Some recent legislation has therefore sought to level the playing field by reducing ecosystems' data advantage. For example, in the European Union, the Digital Markets Act (DMA) prohibits large ecosystems from combining data across markets without explicit user consent, and also facilitates data sharing with smaller competitors via data interoperability.[4] Other countries such as China, India, and the UK, have sought to reduce ecosystems' data advantage via firm-specific regulation.[5] In addition, the emergence of data spaces such as Gaia-X may eventually enable specialized firms to share data with other, and thus compete more effectively with ecosystems.[6] More broadly, structural remedies—such as breaking up big tech companies—that are under consideration by multiple jurisdictions, may also help rein in ecosystems' cross-market data advantage.[7]

However, it is unclear how these policies might affect competition between ecosystems and specialized firms. For example, will specialized firms necessarily innovate more and be better off? How will market prices and the ecosystem's incentive to innovate change? Will these policies ultimately benefit consumers? In this paper we develop a

[3]See, e.g., https://shorturl.at/lDuSH on the importance of data for AI, and https://shorturl.at/J0ln0 on how Gemini 3.0 is likely to leapfrog ChatGPT.

[4]See Articles 5(2) and 6(9) of the legislation https://shorturl.at/pSVZ8.

[5]In China, Ant Group terminated its data-sharing agreement with its former parent company Alibaba in 2022 due to regulatory pressure; see https://shorturl.at/Wd8im. In India, WhatsApp was temporarily banned in 2024 from sharing user data with other entities in the Meta group; see https://shorturl.at/3EIBd. In the UK, under proposed legislation Google would have to allow users to directly port their data to other businesses; see https://shorturl.at/v62ng.

[6]See https://shorturl.at/kHSCN for European initiatives aimed at creating such data spaces.

[7]See, e.g., https://shorturl.at/kjAjP and https://shorturl.at/WIyJp for recent proposals to break up big tech companies like Google.

framework to address these questions.

Our framework features a multiproduct firm competing with multiple single-product rivals across different markets. Firms compete in both price and innovation (quality investment), and the multiproduct firm exhibits cross-product externalities, meaning that the sales it makes on one product affect the surplus that it offers to consumers on other products. In the context of this paper, the multiproduct firm is a digital ecosystem, and the externality arises from cross-market data usage, as discussed above. At the same time, the framework can also apply to traditional industries like retailing and banking. In the former, big-box retailers compete with category specialists and joint brand effects can generate cross-product externalities; in the latter, big financial institutions compete with specialized banks and customer information across business lines may create similar externalities. However, the policy implications of our model are less relevant for these non-digital applications.

Cross-product externalities have an important impact on the way firms price. The ecosystem takes into account that higher sales in one market affect demand and thus profit in other markets—and so it tends to charge lower (resp. higher) prices when cross-product externalities are positive (resp. negative). A single-product firm's problem is more complicated: if it sells more, this affects not only the ecosystem's sales in its own market, but also the ecosystem's sales in all *other* markets—which then feeds back into consumers' demand for its own product. Hence single-product firms also tend to charge lower prices when cross-product externalities are positive (and vice versa), though the underlying mechanism is quite different. On the other hand, in our setup a firm's investment depends only on its sales, so cross-product externalities affect innovation incentives indirectly through their impact on sales. As expected, we face similar technical issues concerning equilibrium existence and uniqueness as does the literature on competition with network effects. However our problem is even more challenging, not least because of the complicated "feedback effect" in single-product firms' problems. Therefore, while we can fully solve the model in some special cases, in general the model is not tractable enough. This motivates us to focus on the case where the number of markets becomes large, and so each single-product firm's feedback effect becomes negligible.

A convenient approach is to consider a continuum version of our model, which under mild conditions can be viewed as the limit of the discrete model with a large number of markets. Since individual markets are atomistic, the feedback effect disappears entirely, and we are able to provide conditions for the existence and uniqueness of an interior equilibrium in which both firms in each market make positive sales.[8] We then use this tractable continuum framework to investigate the impact of three regulatory policies designed to level the playing field between the ecosystem and single-product firms.

Our first policy restricts the ecosystem's ability to use data from one market to improve its products in other markets. This makes the ecosystem's products less attractive, and so tends to benefit the single-product firms. However, consumers are not necessarily better off: first, while single-product firms innovate more, they also set higher prices; second, the ecosystem innovates less, but its prices may not decrease because the policy dampens its incentive to acquire data; third, the regulation directly reduces the ecosystem's ability to use data to improve its products. The policy is more likely to benefit consumers when single-product firms are relatively good at innovating, because in this case their quality investment responds more strongly to the regulation. Innovation is therefore a crucial part of our set-up: if firms compete only in price, the innovation channel disappears and the policy is unambiguously bad for consumers.[9] We also argue that breaking up the ecosystem has qualitatively the same welfare impact as restricting its cross-market data usage.

Our second policy forces the ecosystem to share data with single-product firms.[10] If the ecosystem shares data for free, the impact on prices and innovation is qualitatively the same as under the first policy, but consumers are better off compared to the first policy. Intuitively, both policies make the ecosystem's products relatively less attractive; however, the first one does this by making the ecosystem's products worse, whereas the second one makes the single-product firms' products better. Nevertheless, free data sharing can still harm consumers if single-product firms are relatively poor at innovating. If instead the ecosystem is compensated for its data, it has more incentive to collect data by reducing its prices. We show that there exist compensation schemes

[8]We focus on an interior equilibrium because it is rare that big tech companies completely exclude single-product competitors. Although such firms may sometimes exit, new single-product firms—potentially equipped with better technologies—frequently enter and challenge big tech companies.

[9]We also note that innovation in digital markets is a key concern for policymakers. For instance, "promoting innovation" and ensuring a "high quality of digital products and services" are important objectives of the European Union's Digital Markets Act.

[10]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 for free with new entrants, such as fintech lenders, if consumers consent. (See, e.g., He, Huang, and Zhou, 2023.)

such that single-product firms are willing to buy the ecosystem's data, and relative to no policy intervention consumers are better off due to the lower prices.

Our third policy involves the creation of a "data cooperative," where single-product firms share data with each other but still choose prices and qualities independently. Single-product firms face a trade-off when forming a cooperative: they can use each other's data to improve their products, but the ecosystem prices more aggressively in order to deprive the cooperative of data. We find that a cooperative is more likely to emerge when single-product firms' initial market shares (and so their data endowments) are relatively high. Consumers are not necessarily better off, however, because single-product firms raise their prices and the ecosystem invests less in quality. Nevertheless, similar to our second policy, when single-product firms pay each other for their data, there exist compensation schemes such that a cooperative emerges and consumers are better off compared to no cooperative.

Related literature. Our paper relates to several different strands of literature.

Digital ecosystems. Our paper contributes to an emerging literature in economics on digital ecosystems.[11] Some papers provide a theory for why ecosystems emerge. In Condorelli and Padilla (2024) a firm that is dominant in one market may enter and price aggressively in a secondary market (i.e., form an ecosystem), so as to acquire data that helps foreclose entry into its primary market. Heidhues, Köster, and Kőszegi (2024) assume that, due to default effects, if a consumer buys from a multiproduct firm in the primary market, she is more likely to also buy from that firm in the secondary market. They show that the leader in the primary market can profit from this by acquiring a firm in the secondary market (thus forming an ecosystem). Different from these two papers, we do not consider the formation of an ecosystem. Instead, we focus on how an existing ecosystem influences innovation and expansion by its single-product rivals. We do this using a fully-fledged model of competition (in both price and quality innovation), which is absent from these two papers.

Other papers take the market structure as fixed. In Kraemer and Shekhar (2025)

[11] Platforms like e-commerce websites or app stores are also sometimes considered digital ecosystems. Bisceglia and Tirole (2024) take this perspective and examine issues such as excessive platform fees and self-preferencing by gatekeeper platforms. Anderson and Bedre Defolie (2025) examine the interplay between platform commissions, device fees, and entry decisions of app providers. Jeon, Lefouili, Li, and Simcoe (2025) use a network approach to study optimal pricing by a monopoly supplier of devices that are linked by demand-side externalities.

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 a single-product firm. They find that consumers are always better off when the ecosystem is forced to share primary-market data with its rival.[12] In contrast, we study a richer model with price competition and more general data effects where an ecosystem competes in many markets, each of which can generate and use data. We find that forcing the ecosystem to share data for free can harm consumers by softening competition. We also examine data exchange between single-product firms through a data cooperative, an issue that cannot be studied in their set-up.

Data-driven product improvement. There are other papers that explore data-driven product improvement within a single market.[13] Prüfer and Schottmüller (2021) and Hagiu and Wright (2023) study dynamic duopoly models, in which single-product firms use customer data from previous transactions to help reduce their cost or offer better products. In the former paper firms compete in quality only, and a firm's cost of producing quality decreases in the number of customers it served in the previous period. In the latter paper firms compete in price only, and the value of a firm's product is increasing in the number of its past users. Both papers show that data-driven product improvements tend to lead to market tipping. They also study how policies such as data sharing affect firms' incentives to acquire data, and thus market structure. In contrast, we focus on cross-market data usage by a multiproduct firm that competes with multiple single-product rivals. We allow for both price and quality competition, and we consider various policies that affect how data is leveraged across different markets.

Multiproduct vs single-product firms. Few papers study asymmetric competition between a multiproduct firm and multiple single-product firms. One exception is recent work on conglomerate mergers by Rhodes and Zhou (2019) and Chen and Rey (2023).[14] In both papers, two firms initially operating in separate markets can merge

[12] Bhargava, Dubus, Ronayne, and Shekhar (2026) use a similar model to study whether the single-product firm has an incentive to share its data with the multiproduct firm.

[13] Empirical research indicates 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).

[14] A 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

into a multiproduct firm and then compete with the remaining single-product firms in each market. Such a merger can be profitable due to, respectively, one-stop shopping convenience when consumers face search frictions, or consumption synergies. These papers, however, focus only on price competition, do not have data-driven cross-product externalities, and address very different research questions to us.

Competition with network effects. Our paper is also related to the literature on competition with network effects (see, e.g., Katz and Shapiro, 1985, and the survey by Farrell and Klemperer, 2007). In modeling competition, we adopt the discrete-choice approach as in Chapter 7.8 of Anderson, de Palma, and Thisse (1992). However, the network-effect literature primarily studies single-product competition, whereas we examine multi-product competition and emphasize cross-market network effects. Moreover, we consider asymmetric firms that compete in both price and innovation. These features make it challenging to establish the existence and uniqueness of equilibrium.

Multi-sided platforms. Due to data spillovers across products, the ecosystem in our model can also be regarded as a multi-sided platform. Existing (theoretical) work in that literature, however, usually focuses on *symmetric* competitive platforms for tractability. See, e.g., Armstrong (2006), Rochet and Tirole (2006), Tan and Zhou (2021), and the survey paper by Jullien, Pavan, and Rysman (2021). One exception is Peitz and Sato (2025), who study price competition among two-sided platforms that can differ in their network effects, costs, and exogenously-given qualities; they show that the model is tractable when taste shocks are logistic, and network effects are logarithmic.[15] We study a different type of asymmetric competition between a multiproduct ecosystem and multiple single-product firms. Our focus is also more on the welfare effects of various data policies. Finally, we note that innovation (quality investment) plays an important role in our analysis, but is not usually studied in the literature on multi-sided platforms (including the above papers).

The rest of the paper is organized as follows. We introduce the discrete model in Section 2 and analyze it in Section 3. We then turn to the more tractable continuum version of the model in Section 4 and examine the three regulatory policies. We return

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.

[15]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 asymmetries, since analytical tractability is not a concern. Sometimes they also consider endogenous quality choices (see, e.g., Fan, 2013).

to a tractable special case of the discrete model in Section 5, before concluding in Section 6 with a discussion of possible future avenues for research.

2 The Model

There are $n \geq 2$ product markets which are assumed to be intrinsically independent. Each product $i \in \{1, \cdots, n\}$ 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 investment/innovation.

Let $p_{e,i}$ and $q_{e,i}$ denote the ecosystem's price and quality investment on product i, and let $p_{s,i}$ and $q_{s,i}$ denote single-product firm i's price and quality investment.[16] We assume that the fixed costs of investment for product i are given by respectively

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

where $\eta_{e,i}$ and $\eta_{s,i}$ capture innovation efficiency and can differ across firms and markets. For example, if single-product firm i is more efficient at innovating than the ecosystem in market i due to its specialization, we would have $\eta_{s,i} > \eta_{e,i}$. We assume that both the ecosystem and single-product firms have a constant marginal production cost, 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.)

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).[17] It does not matter for our analysis if the same consumers are present in each market, or if different markets have different consumers. Let $z_i \in [0, 1]$ denote the fraction of consumers who buy product i from the ecosystem, and then $1 - z_i$ is the sales of single-product firm i. Let $\mathbf{z} = \{z_1, \cdots, z_n\}$ denote the ecosystem's sales vector.

Sales generate data, and data is assumed to affect product quality: if a consumer

[16] 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. Under this interpretation, we implicitly assume a linear advertising-disutility function. Allowing for a nonlinear disutility function, however, does not alter the basic logic of our analysis.

[17] Allowing markets to differ in their size does not affect our analysis qualitatively. The assumption of full market coverage, however, is important for tractability.

buys product i from the ecosystem, her surplus is

$$q_{e,i} - p_{e,i} + \delta_{e,i}(\mathbf{z}) + \varepsilon_{e,i},$$

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

$$q_{s,i} - p_{s,i} + \delta_{s,i}(1 - z_i) + \varepsilon_{s,i},$$

where functions $\delta_{e,i}(\mathbf{z})$ and $\delta_{s,i}(1 - z_i)$ capture the effect of data on product quality, and $\varepsilon_{e,i}$ and $\varepsilon_{s,i}$ are taste shocks that capture idiosyncratic consumer preferences. Note that $\delta_{e,i}(\mathbf{z})$ depends on the ecosystem's entire sales vector, so it includes both a *within*-market data effect and a *cross*-market data effect; however, since single-product firm i only has data from its own market, $\delta_{s,i}(1-z_i)$ has only a within-market data effect. Both functions are assumed to be continuously differentiable. As noted in the introduction, our model also admits non-data interpretations and can be applied to any setting with cross-product demand effects.

Given the assumption of full market coverage, only the ecosystem's *relative* data advantage

$$\delta_i(\mathbf{z}) \equiv \delta_{e,i}(\mathbf{z}) - \delta_{s,i}(1 - z_i) \tag{1}$$

and the *relative* taste shock $\varepsilon_{s,i} - \varepsilon_{e,i}$ matter for consumer choices. We assume that for each market i, $\varepsilon_{s,i} - \varepsilon_{e,i}$ 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.[18] 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, in each market there is symmetric product differentiation between the ecosystem and the single-product firm.[19] Note that the log-concavity of f_i implies that both F_i and $1-F_i$ are log-concave. Note also that $f_i(0)$ is the density of indifferent consumers when the two firms in market i offer the same utility. Therefore, a smaller $f_i(0)$ indicates greater product differentiation in market i.

We now make a few remarks and also introduce assumptions on the data effect:

First, our main premise is that more data improves product quality, i.e., $\delta_{e,i}(\mathbf{z})$ and $\delta_{s,i}(1 - z_i)$ are increasing functions. Nonetheless, we also allow for the possibility that

[18]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.

[19]We obtain qualitatively the same results if we allow for asymmetric product differentiation, such that f_i is symmetric around some $\hat{x}_i \neq 0$.

they are decreasing functions. For instance, a firm with more data may face a higher risk of hacking, making it a less secure option for consumers. In non-data interpretations, greater sales can generate congestion effects that also reduce consumer utility. Note also that our data effects may exhibit increasing or decreasing returns to scale, and can accommodate data substitutability or complementarity across markets.

Second, we sometimes consider special cases of the data effect:

(a) Linear data effect:

$$\delta_{e,i}(\mathbf{z}) = w_{e,i}z_i + \sum_{j \neq i} c_{ij}z_j \quad \text{and} \quad \delta_{s,i}(1-z_i) = w_{s,i}(1-z_i),$$

where coefficients $w_{e,i}$ and $w_{s,i}$ measure the *within-market* data effect, and c_{ij} measures the product-pair specific *cross-market* data effect (i.e., the impact of data from product j on product i).

(b) Data effect with a data aggregator:

$$\delta_{e,i}(\mathbf{z}) = \phi_i(A) \quad \text{where} \quad A = \sum_{j=1}^{n} a_j z_j$$

is the data aggregator, $a_j \geq 0$ measures how good product j is at generating data, and $\phi_i(\cdot)$ is a differentiable function which measures how good product i is at using data. Here $\phi_i(\cdot)$ allows for flexible patterns of returns to scale in data-driven quality.

Third, we make the following technical assumption on $\delta_i(\mathbf{z})$ defined in (1):

Assumption 1. *For any i, there exists $\epsilon > 0$ (which can be arbitrarily small) such that*

$$|\delta_i(\mathbf{z}') - \delta_i(\mathbf{z})| \leq \frac{1-\epsilon}{f_i(0)}||\mathbf{z}' - \mathbf{z}||_\infty$$

for any $\mathbf{z}, \mathbf{z}' \in [0,1]^n$, where $||\mathbf{z}||_\infty = \max_i |z_i|$.

Given the differentiability of $\delta_i(\mathbf{z})$, this Lipschitz continuity condition is equivalent to

$$\sum_{j=1}^{n} \left| \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} \right| < \frac{1-\epsilon}{f_i(0)} \tag{2}$$

at any $\mathbf{z}$. In other words, Assumption 1 requires that in each market the total data-driven quality advantage the ecosystem gains when it sells one more unit of each product is bounded by the degree of product differentiation. As we will see, this condition is crucial for ensuring equilibrium uniqueness in our model.

The timing of the game is as follows. Firms first simultaneously choose prices and quality investments in each market. Consumers observe those choices. Following the standard approach in the literature on network effects, consumers form a rational expectation about the sales $\mathbf{z}$ in all markets, and then simultaneously make their purchase decisions.

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

(i) *Two channels of quality improvement.* In our model product quality can be improved via two channels: quality investment and data-driven enhancements. The former reflects efforts such as developing new service features or 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 this 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, as we will see later, having this innovation channel can qualitatively affect the impact of data policies.

(ii) *Data and innovation cost.* An alternative modeling approach for the data effect would be to assume that more data reduces a firm's innovation cost. Such a model can be analyzed in a similar way to the one studied here, and it yields similar insights.

(iii) *Monetization of Data.* Firms may monetize their data by selling it to other firms. The cases in which the ecosystem sells data to single-product firms, or single-product firms sell data to each other, are considered in Section 4.2 as part of our analysis of data-sharing and data-cooperative policies. Selling data to firms outside the model can be incorporated through a straightforward extension. For example, when the ecosystem monetizes its data, it gets additional revenue which is an increasing function of its sales/data; this induces it to cut its prices further, harming single-product firms. However, this extension does not qualitatively alter our analysis or main insights.

(iv) *One-stop-shopping convenience.* If the same set of consumers is present in multiple markets, they may find it more convenient to buy from the ecosystem due to one-stop-shopping benefits. This can be captured by assuming that consumers incur an additional cost when buying from each single-product firm. However, this is equivalent to a leftward shift in the product-differentiation distribution in each market and does not fundamentally affect our analysis.

We have also made some simplifying assumptions such as quadratic cost functions and the presence of only one single-product firm in each market. Relaxing these assumptions does not alter the basic logic of our analysis, but establishing equilibrium existence and uniqueness is more challenging, and comparative statics are less tractable.

3 Analysis

We now solve the model, starting with the consumer problem, before moving to firms' choices of price and quality investment.

3.1 Consumer Problem

We first introduce some notation. Let $v_{e,i} \equiv q_{e,i} - p_{e,i}$ and $v_{s,i} \equiv q_{s,i} - p_{s,i}$ denote the basic surplus of the ecosystem's product i and single-product firm i's product, respectively. Let $v_i \equiv v_{e,i} - v_{s,i}$ be the difference between the two firms' basic surpluses.

For a consumer in market i, if she expects the ecosystem's sales profile to be $\mathbf{z}$, she buys the ecosystem's product i if

$$\varepsilon_{s,i} - \varepsilon_{e,i} \leq v_i + \delta_i(\mathbf{z})$$

and otherwise buys from the single-product firm i. Demand for the ecosystem's product i is therefore $z_i = F_i(v_i + \delta_i(\mathbf{z}))$, where recall that F_i is the CDF of $\varepsilon_{s,i} - \varepsilon_{e,i}$, and demand for single-product firm i is $1 - z_i$. The realized demand must coincide with the expected one in equilibrium. Therefore, for any given profile of prices and quality investments, the equilibrium of the consumer choice game is characterized by the following system of equations: for each $i \in \{1, \cdots, n\}$,

$$z_i = F_i(v_i + \delta_i(\mathbf{z})). \tag{3}$$

As shown in the following lemma, Assumption 1 ensures that the above equations imply a contraction mapping and so must have a unique solution that defines the demand system in our model.

Lemma 1. *Under Assumption 1, for any given prices and qualities, the consumer choice game has a unique equilibrium $\mathbf{z} \in [0,1]^n$, which solves the system of equations in (3), and each z_i increases in v_i. If all $\delta_i(\mathbf{z})$'s are increasing functions, each z_i also increases in v_j for $j \neq i$.*

Proof. Let $\mathcal{F} : [0,1]^n \to [0,1]^n$ denote the mapping defined by the right-hand side of (3). Notice that

$$|F_i(v_i + \delta_i(\mathbf{z})) - F_i(v_i + \delta_i(\mathbf{z}'))| \leq f_i(x)|\delta_i(\mathbf{z}) - \delta_i(\mathbf{z}')|$$
$$\leq f_i(0)|\delta_i(\mathbf{z}) - \delta_i(\mathbf{z}')| \leq (1-\epsilon)||\mathbf{z} - \mathbf{z}'||_\infty,$$

where the first inequality uses the mean value theorem, the second uses single-peakedness of $f_i(x)$ at $x = 0$, and the third uses Assumption 1. Since this holds for any i, we have

$$||\mathcal{F}(\mathbf{z}) - \mathcal{F}(\mathbf{z}')||_\infty \leq (1-\epsilon)||\mathbf{z} - \mathbf{z}'||_\infty.$$

Given the space of $[0,1]^n$ under the maximum norm is complete, the contraction mapping theorem then implies existence and uniqueness of a solution. The proof for the monotonicity of z_i in v_i and v_j is relegated to Appendix A.1. □

Demand "slopes down" for both the ecosystem and the single-product firms: when a firm chooses a lower price or a higher quality for a product, it sells more of that product. If the ecosystem's net data advantage $\delta_i(\mathbf{z})$ in each market is increasing in its sales, it also benefits from demand complementarity: when it sets a lower price or a higher quality on one product, this raises demand for all its other products.

3.2 Firm Problem

We now turn to the firms' problems. As discussed in the introduction, we look for an interior (pure-strategy) equilibrium with $z_i \in (0,1)$ in each market i. Single-product firm i aims to

$$\max_{p_{s,i}, q_{s,i}} \; p_{s,i}(1 - z_i) - \frac{q_{s,i}^2}{2\eta_{s,i}}$$

given the prices and qualities chosen by the ecosystem and other single-product firms. The ecosystem aims to

$$\max_{\mathbf{p}_e, \mathbf{q}_e} \; \sum_{i=1}^{n} \left[p_{e,i} z_i - \frac{q_{e,i}^2}{2\eta_{e,i}} \right]$$

given all single-product firms' choices, where $\mathbf{p}_e$ and $\mathbf{q}_e$ are the ecosystem's price and quality vectors. The complication here is that, since the consumer choice problem has no analytical solution $\mathbf{z}$ in general, each optimization problem is subject to the n constraints in (3).

In the following analysis, for notational convenience, let

$$h_i(z) \equiv \frac{dF_i^{-1}(z)}{dz} = \frac{1}{f_i(F_i^{-1}(z))} \tag{4}$$

denote the *quantile density* function of F_i. Let

$$\tilde{\sigma}_i(x) \equiv 1 + \frac{d}{dx}\left(\frac{1 - F_i(x)}{f_i(x)}\right) = -\frac{[1 - F_i(x)]f_i'(x)}{f_i(x)^2}$$

denote the curvature of $1 - F_i(x)$. Then we define

$$\sigma_i(z) \equiv \tilde{\sigma}_i(F_i^{-1}(z)) = (1 - z)\frac{h_i'(z)}{h_i(z)}, \tag{5}$$

where the last equality can be checked by using the definition of $h_i(z)$. Under our log-concavity condition, $zh_i(z)$ and $(1 - z)h_i(z)$ are respectively increasing and decreasing functions, and the latter implies $(1 - z)h_i'(z) \leq h_i(z)$ and so $\sigma_i(z) \leq 1$.

Ecosystem problem. The ecosystem's problem is relatively easy to deal with if we adopt a *quantity* approach. From (3), we can solve $p_{e,i}$ as

$$p_{e,i} = q_{e,i} - F_i^{-1}(z_i) + \delta_i(\mathbf{z}) - v_{s,i} \tag{6}$$

where recall that $v_{s,i} = q_{s,i} - p_{s,i}$. Given the one-to-one correspondence between $(\mathbf{p}_e, \mathbf{q}_e)$ and $(\mathbf{q}_e, \mathbf{z})$ under Assumption 1, we can then rewrite the ecosystem's problem as an optimization problem with respect to $(\mathbf{q}_e, \mathbf{z})$ without any constraints:

$$\max_{\mathbf{q}_e,\mathbf{z}} \sum_{i=1}^{n} \left\{ \left[q_{e,i} - F_i^{-1}(z_i) + \delta_i(\mathbf{z}) - v_{s,i}\right] z_i - \frac{q_{e,i}^2}{2\eta_{e,i}} \right\}.$$

The first-order condition with respect to $q_{e,i}$ yields

$$q_{e,i} = \eta_{e,i} z_i. \tag{7}$$

When the ecosystem increases $q_{e,i}$ by one unit, the marginal cost is $q_{e,i}/\eta_{e,i}$, while the marginal benefit is z_i since, under our additive utility structure, the ecosystem can raise price by one unit without losing any of its z_i customers on that product.

From the first-order condition with respect to z_i, we can derive

$$p_{e,i} = z_i h_i(z_i) - z_i \frac{\partial \delta_i(\mathbf{z})}{\partial z_i} - \sum_{j \neq i} z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} \tag{8}$$

by using (6). If there were no data effect, raising the sales of product i by one unit would yield a marginal benefit of $p_{e,i}$, but it would also require a price cut of $h_i(z_i)$, resulting in a marginal cost of $z_i h_i(z_i)$. Due to the data effect there are two additional terms in the $p_{e,i}$ expression. To illustrate, consider the case where cross-product spillovers are positive. Then, when the ecosystem sells one more unit of product i, the data-driven surplus that it offers on product j increases by $\frac{\partial \delta_j(\mathbf{z})}{\partial z_i}$; since the ecosystem can raise the price of product j by this amount and not lose any sales, it obtains an additional benefit of $z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i}$. The second term in (8) captures the within-market data effect for $j = i$, and the third term captures the cross-market data effect for $j \neq i$. We call these two terms a "data subsidy," because the ecosystem lowers its prices so as to acquire more data.

It is also convenient to investigate the ecosystem's second-order conditions under this quantity approach.

Lemma 2. *Given single-product firms' prices and qualities, the ecosystem's problem, whenever it has an interior solution, is solved by the first-order conditions (7) and (8) if, for each i,*

$$\eta_{e,i} + \lambda \leq \min_{z \in [0,1]} h_i(z)[2 - \sigma_i(z)] \tag{9}$$

where λ is the largest eigenvalue of the Hessian matrix of $\sum_i z_i \delta_i(\mathbf{z})$ across all $\mathbf{z}$.

Proof. Given any $\mathbf{z}$, the ecosystem's profit is obviously concave in $\mathbf{q}_e$. Using the optimal solution $q_{e,i} = \eta_{e,i} z_i$, we can rewrite its profit as a function of $\mathbf{z}$ only:

$$\sum_i \left[\frac{\eta_{e,i} + \lambda}{2} z_i - F_i^{-1}(z_i) - v_{s,i} \right] z_i + \sum_i \left[z_i \delta_i(\mathbf{z}) - \frac{\lambda}{2} z_i^2 \right].$$

Given the definition of λ, the eigenvalues of the symmetric matrix $\nabla^2 \big(\sum_i z_i \delta_i(\mathbf{z}) \big) - \lambda \mathbf{I}$ are non-positive for any $\mathbf{z}$ and so it is negative semi-definite. Therefore, $\sum_i z_i \delta_i(\mathbf{z}) - \frac{\lambda}{2} \mathbf{z}^T \mathbf{z}$, which equals the second portion of the above profit function, must be concave in $\mathbf{z}$. The first portion is concave in $\mathbf{z}$ if $\eta_{e,i} + \lambda \leq [z_i F_i^{-1}(z_i)]''$ for any $z_i \in [0, 1]$ and for each i, which is equivalent to (9).[20] □

[20]To see that, notice that $[z_i F_i^{-1}(z_i)]'' = 2h_i(z_i) + z_i h_i'(z_i)$, so $\min_{z_i \in [0,1]} [z_i F_i^{-1}(z_i)]'' = \min_{z_i \in [0,1]} [2h_i(1 - z_i) + (1 - z_i) h_i'(1 - z_i)]$ as z_i and $1 - z_i$ share the same range. The symmetry of f_i implies $h_i(1 - z_i) = h_i(z_i)$ and $h_i'(1 - z_i) = -h_i'(z_i)$, and therefore the latter expression equals $2h_i(z_i) - (1 - z_i) h_i'(z_i) = h_i(z_i)[2 - \sigma_i(z_i)]$ by using the definition of σ_i in (5).

Given $\sigma_i(z) \leq 1$ and the single-peakedness of f_i, the right-hand side of (9) is greater than $1/f_i(0)$. Hence, (9) must hold if the degree of product differentiation is sufficiently large relative to innovation efficiency and the data effect. Note that λ summarizes the sharpest change in the slope of $\sum_i z_i \delta_i(\mathbf{z})$, so it controls the strength of the data effect when the ecosystem gains more data. To illustrate, it is helpful to see an example where λ can be computed explicitly. Suppose that $\delta_i(\mathbf{z}) = b_i A$, where $A = \sum_j a_j z_j$ is a data aggregator and we have assumed $\delta_{s,i}(1 - z_i) = 0$ for each market i. (Here a_i measures how good product i is at generating data and b_i measures how good it is at using data.) Then $\sum_i z_i \delta_i(\mathbf{z}) = \sum_i a_i z_i \sum_i b_i z_i$, and its Hessian is $[a_i b_j + b_i a_j]_{i,j \in \{1,\cdots,n\}}$. This special symmetric matrix has a largest eigenvalue $\lambda = \sum_i a_i b_i + \sqrt{\sum_i a_i^2 \sum_i b_i^2}$, which is clearly larger when products generate or use more data.

Single-product firm problem. A single-product firm's problem is more complicated: it only controls its own price and quality investment but faces n constraints, so the quantity approach is not helpful here. (When single-product firm i adjusts its price or investment, this affects sales in *every* market through the system of demand equations in (3), thereby generating the n constraints.) It is now more convenient to use the Lagrangian method. Single-product firm i's problem is:

$$\max_{(p_{s,i}, q_{s,i}, \mathbf{z})} \; p_{s,i}(1 - z_i) - \frac{q_{s,i}^2}{2\eta_{s,i}} + \sum_{j=1}^{n} \mu_j^i \left[F_j^{-1}(z_j) - v_j - \delta_j(\mathbf{z}) \right]$$

where $\{\mu_j^i\}_{j=1}^n$ are the firm i specific Lagrangian multipliers, and where μ_j^i is the shadow value to single-product firm i of reducing the ecosystem's relative data advantage in market j by one unit.

From the first-order conditions with respect to $(p_{s,i}, q_{s,i})$, we derive $\mu_i^i = 1 - z_i$ and

$$q_{s,i} = \eta_{s,i}(1 - z_i). \tag{10}$$

The explanation for the optimal quality investment is the same as in the ecosystem's problem. From the first-order conditions with respect to z_i and using $\mu_i^i = 1 - z_i$, we derive

$$p_{s,i} = (1 - z_i) h_i(z_i) - (1 - z_i) \frac{\partial \delta_i(\mathbf{z})}{\partial z_i} - \sum_{j \neq i} \mu_j^i \frac{\partial \delta_j(\mathbf{z})}{\partial z_i}. \tag{11}$$

The explanation for the $(1 - z_i) h_i(z_i)$ term is the same as in the ecosystem's problem. There are also two data effects here: the first is a standard within-market data effect on pricing, while the second captures a "feedback effect" due to cross-market data spillovers within the ecosystem. Suppose, for example, that each $\delta_i(\mathbf{z})$ is increasing in $\mathbf{z}$. When single-product firm i sells more, it reduces the ecosystem's data advantage not only in market i but also in other markets through the $\delta_j(\mathbf{z})$'s. This data reduction in other markets feeds back into market i, further eroding the ecosystem's data advantage there and yielding an extra benefit for single-product firm i.

We can pin down the other Lagrangian multipliers $\{\mu^i_{j\neq i}\}$ from the first-order conditions with respect to $\{z_{j\neq i}\}$:

$$\mu^i_j h_j(z_j) = \sum_{l=1}^{n} \mu^i_l \frac{\partial \delta_l(\mathbf{z})}{\partial z_j}. \tag{12}$$

As detailed in the Online Appendix B.1.1, $\mu^i_j(\mathbf{z}) = (1 - z_i)\nu^i_j(\mathbf{z})$, where $1 - z_i$ is simply because reducing the ecosystem's data advantage in market j allows single-product firm i to increase its price on each unit of its sales, while $\nu^i_j(\mathbf{z})$ does not have a simple explicit expression, except in some special cases.[21]

Another challenge in the single-product firms' problem concerns their second-order conditions:

Lemma 3. *Given other firms' prices and qualities, single-product firm i's problem, whenever it has an interior solution, is solved by the first-order conditions if* $1 - z_i$ *is log-concave in* v_i *and* $\eta_{s,i}\frac{\partial z_i}{\partial v_i} \leq 1$*. If all* $\delta_i(\mathbf{z})$*'s and* $F_i(x)$*'s are linear, a sufficient condition is* $\eta_{s,i}\frac{\partial z_i}{\partial v_i} \leq 2$*.*

Unfortunately, as detailed in the proof in the Online Appendix B.1.2, it is hard to find more primitive conditions for this lemma to hold. One exception, as we will examine more closely in Section 5, is the "symmetric double-linear" case with symmetric product markets, linear data effects, and a linear Hotelling distribution in each market.

3.3 Equilibrium

Substituting the expressions for prices and qualities as functions of $\mathbf{z}$ into $F_i^{-1}(z_i) = v_i + \delta_i(\mathbf{z})$ yields

$$G_i(z_i) = \sum_{j=1}^{n} [z_j - \mu^i_j(\mathbf{z})] \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} + \delta_i(\mathbf{z}), \tag{13}$$

[21] For example, in the two-product case, one can derive that $\nu^i_j(\mathbf{z}) = \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} \Big/ \left[h_j(z_j) - \frac{\partial \delta_j(\mathbf{z})}{\partial z_j}\right]$. However, even with this explicit expression, it remains difficult to solve the single-product firms' second-order conditions and to establish equilibrium existence and uniqueness.

where

$$G_i(z_i) = F_i^{-1}(z_i) + \eta_{s,i} - 2\bar{\eta}_i z_i + (2z_i - 1)h_i(z_i) \tag{14}$$

with $\bar{\eta}_i \equiv (\eta_{e,i} + \eta_{s,i})/2$ includes all terms which do not directly include the data effect, while the right-hand side of (13) is the data-driven relative benefit of buying ecosystem's product i (which includes both firms' data effect on their prices and their direct data-driven quality). This system of equations in $\{z_i\}$ defines the equilibrium sales in each market. Once we solve $\{z_i\}$, we can substitute back into our derived expressions for $(p_{e,i}, q_{e,i}; p_{s,i}, q_{s,i})$ and thus determine all the equilibrium prices and quality investments.

Unfortunately, due to the complexity of the $\mu_j^i(\mathbf{z})$'s and the single-product firms' second-order conditions, in general it is hard to establish conditions for existence and uniqueness of equilibrium, and it is also hard to conduct comparative statics.

To make further progress, in the next section we first study a limit case with a continuum of product markets. (As explained in the Online Appendix B.1.3, the continuum case is essentially the same as taking $n \to \infty$ under a condition ensuring Assumption 1 in the limit.) In the continuum case, we also explore the policy implications of our model. Finally, we return to a tractable discrete case in Section 5.

4 The Continuum Case and Policy Implications

Consider the same setting as above, except that now there is a continuum of markets with unit measure, indexed by $i \in [0,1] \equiv \mathcal{I}$. The ecosystem's sales vector is now $\mathbf{z} = \{z_i\}_{i \in \mathcal{I}}$, a (measurable) function with both domain and range being $[0,1]$. Let $\mathcal{Z}$ denote the space of such functions. The ecosystem's relative data (dis)advantage in market i is now a functional $\delta_i(\mathbf{z})$. We equip $\mathcal{Z}$ with the L^2 norm,[22] and assume that the functional $\delta_i(\mathbf{z})$ is *Fréchet differentiable* at any $\mathbf{z}$. Then, as explained in the Online Appendix B.2.1, the usual partial derivative $\frac{\partial \delta_i(\mathbf{z})}{\partial z_j}$ for any i and j is well defined.[23] The continuum version of Assumption 1 is then equivalent to

$$\int_0^1 \left| \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} \right| dj < \frac{1-\epsilon}{f_i(0)} \tag{15}$$

[22]That is, we focus on equivalent classes of square-integrable functions on $[0,1]$, i.e., those satisfying $||\mathbf{z}||_2 = \left(\int_0^1 z_i^2 di\right)^{1/2} < \infty$. Note that $\mathbf{z}$ and $\mathbf{z}'$ are equal if they differ only on a zero-measure set of i.

[23]Note that $\delta_i(z_j', \mathbf{z}_{-j}) = \delta_i(z_j, \mathbf{z}_{-j})$ in our continuum framework, but this is not inconsistent with $\frac{\partial \delta_i(\mathbf{z})}{\partial z_j} \neq 0$, because the impact of a change in z_j only on $\delta_i(\mathbf{z})$ equals $\frac{\partial \delta_i(\mathbf{z})}{\partial z_j}$ multiplied by the measure of j (which is zero).

for each i. From now on we drop the integral limits when no confusion arises. Note also that in this continuum framework, the amount of data from a single market is negligible compared to that from a positive measure of markets. Hence, $\delta_{s,i}(1-z_i)=0$ and $\delta_i(\mathbf{z})=\delta_{e,i}(\mathbf{z})$. In what follows, for expositional brevity, we abstract from standard zero-measure issues and require the equilibrium analysis to hold in each market i.

Remark. Even in this continuum setting, we could incorporate the within-market data effect by assuming that it is comparable in magnitude to the aggregate cross-market data effect. For instance, data from, say, market i can be much more useful for product i than cross-market data from any other market. We study this case in the Online Appendix B.2.5 and show that our main results remain qualitatively the same.

4.1 Equilibrium

The analysis of the consumer choice game naturally extends to this continuum case, except that now a change in v_i alone has no impact on $z_{j\neq i}$. This implies that a single-product firm's price or quality change does not affect sales in any other market.[24] As a result, the "feedback effect"—the main complication in the discrete case—disappears.

The analysis of the ecosystem's problem remains unchanged, including the second-order condition in Lemma 2.[25] However a single-product firm's problem is significantly simpler. Following a similar logic as in the discrete case, we show in the Online Appendix B.2.2 that each firm's quality investment takes the same form as before:

$$q_{e,i}=\eta_{e,i}z_i \quad \text{and} \quad q_{s,i}=\eta_{s,i}(1-z_i) \tag{16}$$

and their optimal prices are respectively:

$$p_{e,i}=z_ih_i(z_i)-\int z_j\frac{\partial\delta_j(\mathbf{z})}{\partial z_i}dj \quad \text{and} \quad p_{s,i}=(1-z_i)h_i(z_i). \tag{17}$$

[24]More formally, suppose that single-product firm i unilaterally changes v_i to v_i'. The system of demand equations becomes $z_i'=F_i(v_i'+\delta_i(z_i',\mathbf{z}_{-i}))$ and $z_j=F_j(v_j+\delta_j(z_i',\mathbf{z}_{-i}))$ for $j\neq i$. Since $\delta_i(z_i',\mathbf{z}_{-i})=\delta_i(\mathbf{z})$ in our continuum setup, $\mathbf{z}_{-i}$ remains unchanged, while $z_i'=F_i(v_i'+\delta_i(\mathbf{z}))$.

[25]In the continuum case we redefine λ in Lemma 2 as:

$$\lambda=\min\left\{\tilde{\lambda}:\int\int\frac{\partial^2\theta(\mathbf{z})}{\partial z_i\partial z_j}x_ix_jdidj\leq\tilde{\lambda}\int x_i^2di \text{ for any } \mathbf{z}\in\mathcal{Z} \text{ and } \mathbf{x}\in\mathbb{R}^{[0,1]}\right\}$$

where $\theta(\mathbf{z})\equiv\int z_i\delta_i(\mathbf{z})di$. That is, λ is the smallest $\tilde{\lambda}$ that makes $\theta(\mathbf{z})-\frac{\tilde{\lambda}}{2}\int z_i^2di$ concave in $\mathbf{z}$, which is what we need in the proof of Lemma 2. This coincides with the largest eigenvalue λ of the Hessian of $\theta(\mathbf{z})$ whenever it exists. Note also that $\lambda\geq 0$ when the $\delta_i(\mathbf{z})$ are increasing functions.

Compared to the discrete case, $p_{e,i}$ has the same expression, though the within-market effect is now negligible, while $p_{s,i}$ is much simpler as both the within-market effect and the feedback effect vanish. Moreover, a primitive second-order condition for single-product firm i is

$$\eta_{s,i} \leq \min_{z\in[0,1]} h_i(z)[2-\sigma_i(z)]. \tag{18}$$

As explained before, the right-hand side is greater than $1/f_i(0)$, so this condition must hold if there is enough product differentiation.

Finally, the system of equations for equilibrium sales simplifies to: for each $i \in [0,1]$,

$$G_i(z_i) = \int z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj + \delta_i(\mathbf{z}) \equiv \Delta_i(\mathbf{z}), \tag{19}$$

where $\Delta_i(\mathbf{z})$ is the data-driven benefit of buying the ecosystem's product i, i.e., the sum of the ecosystem's data subsidy and relative data advantage on product i.

We can derive primitive conditions for the system of equations (19) to have a unique interior solution. Together with the second-order conditions derived before, we have the following result concerning equilibrium existence and uniqueness:

Proposition 1. *In the continuum case, suppose that Assumption 1 (or equivalently (15)) and firms' second-order conditions (9) and (18) hold. If, for each* $i \in [0,1]$*,*

$$\max\{\eta_{s,i} - \underline{\Delta}_i, \eta_{e,i} + \overline{\Delta}_i\} < l_i + \frac{1}{f_i(l_i)} \tag{20}$$

and

$$2\bar{\eta}_i + \lambda < \frac{3}{f_i(0)} \tag{21}$$

where $\underline{\Delta}_i = \min_{\mathbf{z}} \Delta_i(\mathbf{z})$ *and* $\overline{\Delta}_i = \max_{\mathbf{z}} \Delta_i(\mathbf{z})$*, then there exists a unique interior equilibrium, and it is characterized by (16), (17), and (19).*

Condition (20) ensures that the system of equations in (19) have interior solutions, and condition (21) ensures that the solution must be unique. In the example with $\delta_i(\mathbf{z}) = b_i \int a_j z_j dj$, if all the b_i's are positive, we have $\underline{\Delta}_i = 0$ and $\overline{\Delta}_i = a_i \int b_j dj + b_i \int a_j dj$. Note that (20) must hold for unbounded preference distributions with $l_i = \infty$, and a sufficient condition can be obtained by replacing the right-hand side by $3/(2f_i(0))$.[26]

The conditions for existence and uniqueness of equilibrium require that innovation efficiencies and data spillovers not be too large relative to $1/f_i(0)$ (i.e., the level of

[26] This follows because the symmetry of $f_i(\cdot)$ implies $f_i(l_i) \leq f_i(0)$ and $l_i \geq 1/(2f_i(0))$.

product differentiation) in each market. Consider what might go wrong if this were not the case. First, multiple solutions to (19) might exist: if the ecosystem were suddenly expected to sell more, its investment and data-driven quality would increase a lot, while single-product firms' investment would decrease a lot. With low product differentiation, this could lead to enough extra sales for the ecosystem to rationalize the initial change in expectations. Second, these self-fulfilling expectations could even lead to corner cases, where the solution to (19) is such that (in some markets) only one firm makes positive sales. Third, global deviations may be profitable: if the ecosystem offers much lower prices on some products, due to strong cross-product spillovers the demand for all its products could go up by so much that its profit increases.[27]

Double-Linear Example. In general the continuum model has no closed-form solution. An exception is when spillovers and product differentiation are both linear. In particular, suppose that spillovers are $\delta_i(\mathbf{z}) = b_i \int a_j z_j dj$ with $b_i \geq 0$ for all i, and that competition in each market is linear Hotelling, i.e., $F_i(x) = (l_i + x)/(2l_i)$. (Here a larger l_i means higher product differentiation.) The conditions in Proposition 1 for equilibrium existence and uniqueness then simplify to

$$\max\{\eta_{s,i}, \eta_{e,i} + \lambda\} < 4l_i, \quad 2\bar{\eta}_i + \lambda < 6l_i, \quad \max\{\eta_{s,i}, \eta_{e,i} + a_i\mathbb{E}[b] + b_i\mathbb{E}[a]\} < 3l_i,$$

where we use the notation $\mathbb{E}[x] \equiv \int x_i di$ and $\lambda = \mathbb{E}[ab] + \sqrt{\mathbb{E}[a^2]\mathbb{E}[b^2]}$.[28] The equilibrium sales equation (19) becomes $G_i(z_i) = a_i B + b_i A$, where $A = \int a_j z_j dj$ and $B = \int b_j z_j dj$, from which we solve

$$z_i = r_i + \frac{a_i B + b_i A}{g_i},$$

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. Then (A, B) solve the system of linear

[27] In the Online Appendix B.2.3 we also examine how prices and quality investments differ across markets with heterogeneous $\delta_i(\mathbf{z})$. Most results are as expected, except that the ecosystem may charge a *higher* price in markets that generate *more* data (i.e., create higher spillovers to other markets). This is because the ecosystem wishes to sell more in such markets, but when its $\eta_{e,i}$'s are high enough it is better to achieve this through higher investment, allowing it to charge a higher price.

[28] In this example a sufficient condition for a unique equilibrium in the consumer choice game is $\mathbb{E}[ab] < 1/\max_i f_i(0)$, which is weaker than condition (15), and is implied by our conditions here—because the first one implies $\lambda < 2/f_i(0)$, while the Cauchy-Schwartz inequality implies $\lambda \geq 2\mathbb{E}[ab]$.

equations $A = \int a_i z_i di$ and $B = \int b_i z_i di$, which we can write as

$$\begin{bmatrix} 1 - \mathbb{E}\left[ab/g\right] & -\mathbb{E}[a^2/g] \\ -\mathbb{E}[b^2/g] & 1 - \mathbb{E}[ab/g] \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \mathbb{E}[ar] \\ \mathbb{E}[br] \end{bmatrix}.$$

We can then substitute the resulting expressions for A and B into the z_i equation above, and then from earlier obtain explicit expressions for all prices and qualities.

4.2 Data Policies

We now use our continuum framework to evaluate the impact of three policies. The first policy restricts the ecosystem's ability to leverage data across products, the second mandates the ecosystem to share data with specialized firms (possibly with compensation), and the third studies the formation of a data cooperative among specialized firms which allows them to share data with each other (also possibly with compensation). To make the analysis interesting, we assume in this section that data-driven cross-product externalities are positive, so that these policies level the playing field between the ecosystem and single-product firms.

4.2.1 Restricting cross-market data usage

First consider a policy that restricts the ecosystem's ability to use cross-market data, leading to a reduction in its data-driven advantage $\Delta_i(\mathbf{z})$ defined in (19) in a positive measure of markets. Suppose Proposition 1 holds both pre and post regulation, and let $\mathbf{z}^*$ and $\tilde{\mathbf{z}}^*$ denote the corresponding equilibrium sales. We provide conditions such that this policy is unambiguously good for *all* single-product firms. At the same time, we also show that the policy does not necessarily benefit consumers.

Proposition 2. *Consider regulation that reduces $\delta_i(\mathbf{z})$, and leads to strictly lower $\Delta_i(\mathbf{z})$ at the pre-regulation equilibrium sales $\mathbf{z}^*$, in a positive measure of markets. Suppose all $\Delta_i(\mathbf{z})$'s are strictly increasing functions both before and after regulation. Then the solution to (19) decreases strictly from z_i^* to $\tilde{z}_i^*$ in every market i. As a result,*
(i) each single-product firm sells and invests more, charges more, and earns more profit;
(ii) the ecosystem sells and invests less in every market, but it may set higher prices and earn more profit;
(iii) consumer surplus in market i increases if single-product firm i is sufficiently efficient at innovating in terms of

$$\eta_{s,i} > [1 + z_i - \sigma_i(z_i)]h_i(z_i) \tag{22}$$

for all $z_i \in [\tilde{z}_i^*, z_i^*]$. (Consumer surplus decreases if the opposite condition holds.)

Intuitively, the policy weakens the ecosystem's offering in the affected markets, causing it to sell less and collect less data in those markets; due to the cross-market spillovers, the ecosystem then ends up selling and innovating less in *all* markets, including those not directly affected by the policy. Since the regulation shifts demand towards the single-product firms, they optimally raise their prices and invest more.[29] They also earn higher profit. To see this, note that we can write single-product firm i's profit as

$$\pi_{s,i} = (1 - z_i)^2 \left[h_i(z_i) - \frac{\eta_{s,i}}{2} \right]. \tag{23}$$

Using $\sigma_i(\cdot)$ defined in (5), one can check that this profit increases in $1 - z_i$ if and only if $\eta_{s,i} < h_i(z_i)[2 - \sigma_i(z_i)]$, which must hold given the second-order condition (18).

Although the ecosystem sells and innovates less in every market under the regulation, it does not necessarily charge lower prices and make less profit. To see this, first note that ecosystem profit can be written as

$$\Pi_e = \int z_i^2 \left[h_i(z_i) - \frac{\eta_{e,i}}{2} \right] di - \int \int z_i z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj di. \tag{24}$$

One can check that regulation decreases the first part if $\eta_{e,i} < h_i(1 - z_i)[2 - \sigma_i(1 - z_i)]$ for each i, which is implied by the second-order condition (9) (given that $\lambda \geq 0$ under positive data spillovers). However regulation can also decrease the second part, which captures the total cost to the ecosystem of subsidizing data collection. In general either effect can dominate, as we show below in an example. Intuitively, data regulation tends to reduce the ecosystem's incentive to collect data, and this can induce it to raise its prices sufficiently strongly that its profit actually increases. (However, it appears difficult to derive clean primitive conditions under which one effect dominates the other.)

Remark. It may seem surprising that the policy can benefit the ecosystem. However, recall that our analysis starts with an arbitrary collection of $\delta_i(\mathbf{z})$'s. Proposition 2 implies that if the ecosystem could credibly commit to the $\delta_i(\mathbf{z})$'s upfront (which of course may not always be possible), it might choose not to collect or use as much data as it feasibly could. In that case, restricting the ecosystem's cross-market data usage would obviously reduce its profit—but the impact on single-product firms and consumers would be exactly the same as in parts (i) and (iii) of the proposition.

[29]Given our log-concavity condition $q_{s,i} = \eta_{s,i}(1 - z_i)$ and $p_{s,i} = (1 - z_i)h_i(z_i)$ increase in $1 - z_i$.

The regulation also has an ambiguous impact on consumers. Consumer surplus in market i is $V_i = \mathbb{E}[\max\{v_{e,i} + \delta_i(\mathbf{z}) + \varepsilon_{e,i}, v_{s,i} + \varepsilon_{s,i}\}]$. Using $F_i^{-1}(z_i) = v_{e,i} - v_{s,i} + \delta_i(\mathbf{z})$ and $v_{s,i} = q_{s,i} - p_{s,i}$, this can be rewritten as

$$V_i - \mathbb{E}[\max\{\varepsilon_{e,i}, \varepsilon_{s,i}\}] = q_{s,i} - p_{s,i} + \int_0^{F_i^{-1}(z_i)} F_i(x)dx. \tag{25}$$

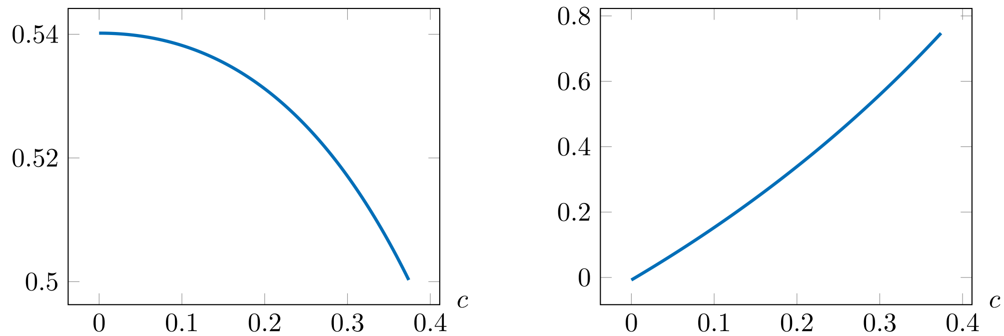
(This expression is valid regardless of the sign of $F_i^{-1}(z_i)$. Note that the data effect $\delta_i(\mathbf{z})$ only affects V_i indirectly through its impact on z_i.) Data regulation affects V_i in three ways. First, it benefits consumers by raising single-product firm i's quality investment $q_{s,i}$. Second though, it harms consumers by inducing that firm to also charge a higher price $p_{s,i}$.[30] Third, it harms consumers by reducing the ecosystem's data-driven quality, and by inducing the ecosystem to invest less and potentially charge more. This is captured by the integral term in (25). Consequently, data regulation *only* benefits consumers when it induces a sufficiently large increase in single-product firms' innovation—meaning that their innovation efficiency $\eta_{s,i}$ must be sufficiently large, as indicated by condition (22).[31] (This also implies that if $\eta_{e,i} = \eta_{s,i} = 0$, such that our model reduces to one of pure price competition, regulation is unambiguously bad for consumers.)

To illustrate the above results, consider a symmetric double-linear example where, in each market, $\delta_i(\mathbf{z}) = c \int z_j dj$ with $c > 0$ being the cross-market data effect. Consider a policy that reduces c. One can check that all conditions in Proposition 2 are satisfied, so the policy unambiguously benefits the single-product firms. However the impact on the ecosystem and consumers depends on parameters. Figure 2 depicts an example where the ecosystem is much more efficient at innovation than single-product firms. A reduction in c leads to a relatively small increase in single-product firms' innovation, but sufficiently softens price competition, so that the ecosystem is better off while consumers are worse off. In contrast, Figure 3 depicts an example in which the single-product firms are much more efficient at innovation than the ecosystem. A reduction in c leads to a relatively large increase in single-product firms' innovation, and so the ecosystem is

[30] Regulation has an ambiguous impact on single-product firm i's basic surplus $v_{s,i} = q_{s,i} - p_{s,i} = (1 - z_i)[\eta_{s,i} - h_i(z_i)]$. This is because it increases in z_i if and only if $\eta_{s,i} < h_i(z_i)[1 - \sigma_i(z_i)]$, but this is not implied by the single-product firm's second-order condition.

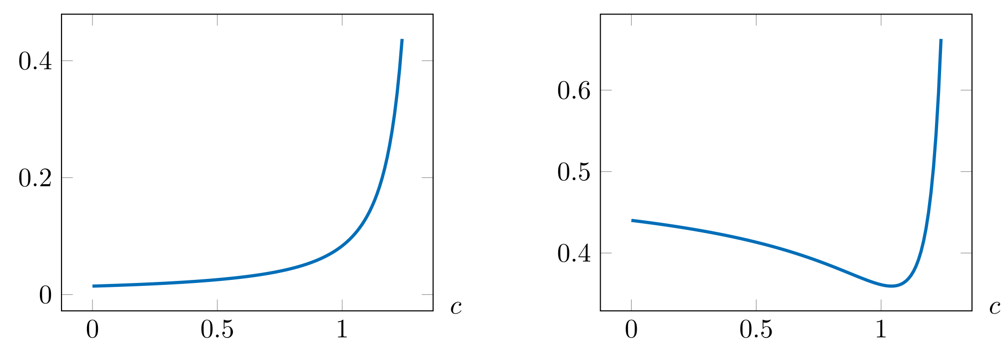
[31] We note that part (iii) of Proposition 2 does not give a primitive condition as z_i is endogenous. In specific examples (e.g., the double-linear one) where z_i is solvable, we can derive an explicit threshold for $\eta_{s,i}$ in terms of primitive parameters.

worse off. Consumers benefit from the increase in single-product firms' innovation, but are harmed by the reduction in the data-driven component of the ecosystem's surplus offering. The first effect dominates when c is not too large.



(a) Ecosystem Profit (b) Consumer Surplus

Figure 2: The impact of restricting cross-market data usage
(Products are symmetric, F is uniform on $[-1, 1]$, $\delta(\mathbf{z}) = c \int z_j dj$, $\eta_e = 2.25$, and $\eta_s = 0.25$.)



(a) Ecosystem Profit (b) Consumer Surplus

Figure 3: The impact of restricting cross-market data usage
(Products are symmetric, F is uniform on $[-1, 1]$, $\delta(\mathbf{z}) = c \int z_j dj$, $\eta_e = 0.5$, and $\eta_s = 2.75$.)

Remark. Here we discuss the conditions required in Proposition 2. For a reduction in $\delta_i(\mathbf{z})$ to lower $\Delta_i(\mathbf{z})$ defined in (19), it is sufficient that the $\delta_i(\mathbf{z})$ are increasing functions that are "scaled down" by the regulation; for $\Delta_i(\mathbf{z})$ to be increasing, it is sufficient that the $\delta_i(\mathbf{z})$ are increasing and weakly convex. Both conditions are satisfied, for example, when the data effect is positive and linear. However, when the conditions fail, as illustrated in the Online Appendix B.2.4 it is possible that data regulation harms single-product firms by inducing the ecosystem to price more aggressively.

Breaking up the ecosystem. Consider a policy that breaks the ecosystem up into two smaller sub-ecosystems. Since each sub-ecosystem has access to less cross-market data, breaking up the ecosystem is qualitatively similar to restricting its cross-market data usage. For instance, consider the case with symmetric markets, and suppose we break the ecosystem up into two units of equal size. This is equivalent to the original large ecosystem having a smaller data effect $\delta_i(\hat{\mathbf{z}})$ with $\hat{z}_i = z_i$ for $i \in [0, 1/2]$ and $\hat{z}_i = 0$ otherwise. The above analysis then implies that breaking up an ecosystem induces the two sub-ecosystems to invest less than the ecosystem did. It also induces the single-product firms to invest more. In addition, this increase in single-product firms' investment must be relatively large in order for the policy to benefit consumers.

4.2.2 Data sharing

We now consider a policy that requires the ecosystem to share data with the single-product firms. A consumer who buys single-product firm i's product now gets surplus

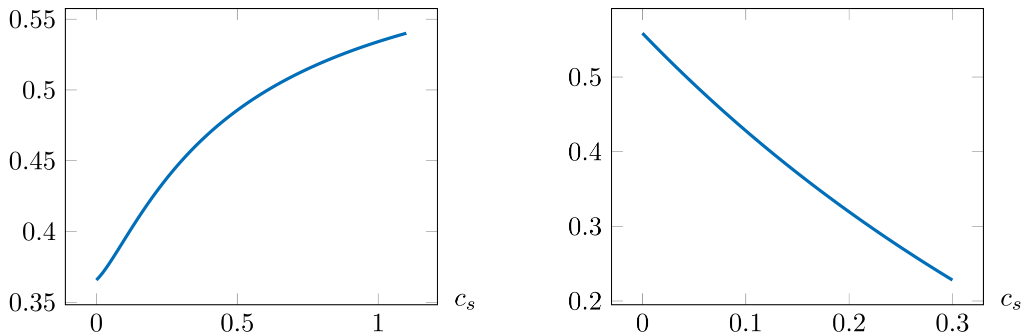
$$q_{s,i} - p_{s,i} + \delta_{s,i}(\mathbf{z}) + \varepsilon_{s,i},$$

where $\delta_{s,i}(\mathbf{z})$ is a positive and increasing functional that captures the utility boost from shared data. (The surplus from buying ecosystem product i is unchanged.)

We distinguish between whether the ecosystem shares its data for free, or is compensated for it. In the latter case, the policymaker publicly chooses a compensation scheme $\{t_i(\mathbf{z})\}_{i \in \mathcal{I}}$, where $t_i(\mathbf{z})$ is Fréchet differentiable and denotes how much single-product firm i must pay the ecosystem to access its data when its sales are $\mathbf{z}$. The timing of this extended game is as follows: each single-product firm simultaneously and publicly decides whether to accept the compensation scheme and buy the ecosystem's data; all firms set their prices and qualities, and then consumers make their choices; finally, data compensation is paid according to realized sales. We find that:

Proposition 3. *(i) Free data sharing has the same impact on firms as restricting cross-market data usage within the ecosystem, but it is strictly better for consumers.*
(ii) There exist data-sharing compensation schemes such that all single-product firms buy the ecosystem's data, and consumer surplus is strictly higher relative to no data sharing.

Start with the case of free data sharing in part (i). Since a single-product firm's behavior still has no impact on $\mathbf{z}$, the equilibrium analysis remains the same as before,



(a) Case where η_s is large relative to η_e (b) Case where η_e is large relative to η_s

Figure 4: The impact of free data sharing on consumer surplus

(Products are symmetric, F is uniform on $[-1, 1]$, $\delta_{e,i}(\mathbf{z}) = c_e \int z_j dj$, and $\delta_{s,i}(\mathbf{z}) = c_s \int z_j dj$. In the left panel $\eta_e = 0.5$, $\eta_s = 2.75$, and $c_e = 1.1$; in the right panel $\eta_e = 2.25$, $\eta_s = 0.25$, and $c_e = 0.3$.)

except that now the ecosystem's net data (dis)advantage is $\delta_i(\mathbf{z}) = \delta_{e,i}(\mathbf{z}) - \delta_{s,i}(\mathbf{z})$.[32] Notice that increasing $\delta_{s,i}(\mathbf{z})$ is equivalent to decreasing $\delta_{e,i}(\mathbf{z})$ when they change $\delta_i(\mathbf{z})$ in the same way. As a result, free data sharing has the same impact on sales, prices, investments and profits as restricting cross-market data usage. However free data sharing is more beneficial (or less harmful) for consumers, because it enables single-product firms to offer higher data-driven quality, whereas the first policy reduces the ecosystem's data-driven quality (and thus the surplus it offers).[33]

Nevertheless, free data sharing can still harm consumers by dampening the ecosystem's incentive to collect data, causing it to sell and innovate less. Figure 4 illustrates this for a symmetric double-linear example where, in each market, $\delta_{e,i}(\mathbf{z}) = c_e \int z_j dj$ and $\delta_{s,i}(\mathbf{z}) = c_s \int z_j dj$. In the left panel, the single-product firms are much more efficient at innovating compared to the ecosystem; a policy that raises c_s benefits consumers, as it induces a relatively large increase in single-product firms' innovation. However, in the right panel, the ecosystem is much more efficient at innovating compared to the single-product firms; a policy that raises c_s harms consumers, as the induced decrease in the ecosystem's innovation dominates.

Part (ii) of Proposition 3 then shows that an appropriately-designed compensation scheme restores the ecosystem's incentive to collect data, raising its sales and its in-

[32]Note that now $\delta_i(\mathbf{z})$ could be decreasing and even negative if single-product firms use data more effectively than the ecosystem. This is another reason why it is useful to allow $\delta_i(\mathbf{z})$ to be decreasing in the baseline model.

[33]More formally, one can check that with data sharing V_i in equation (25) is larger by $\delta_{s,i}(\mathbf{z})$.

vestment, thereby ensuring that data sharing benefits consumers. Consider a general compensation scheme where no single-product firm's choice affects its payment to the ecosystem. The equilibrium analysis is then the same as earlier except that now

$$p_{e,i} = z_i h_i(z_i) - \int z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj - \int \frac{\partial t_j(\mathbf{z})}{\partial z_i} dj. \tag{26}$$

Therefore, when the Fréchet derivatives $\frac{\partial t_j(\mathbf{z})}{\partial z_i}$ are positive, the ecosystem has more incentive to cut prices than before, because any extra data that it collects now generates additional revenues. Next, consider the particular compensation scheme:

$$t_i(\mathbf{z}) = k_i + \int z_j \delta_{s,j}(\mathbf{z}) dj, \tag{27}$$

where k_i is a market-specific fixed fee that can be negative; note that $t_i(\mathbf{z}) - k_i$ is the same across markets, and is unaffected by any single-product firm's choices. We show in the proof that, for appropriately-chosen k_i, all single-product firms accept the compensation scheme in (27) and make positive payments to the ecosystem. We also show that (27) ensures that equilibrium sales $\mathbf{z}$ are the *same* as without any data sharing. This means that all quality investments, as well as all single-product firms' prices, are also the same as without any data sharing. However, the shared data enables single-product firm i to offer $\delta_{s,i}(\mathbf{z})$ more surplus, while the scheme in (27) induces the ecosystem to cut its price in market i, and thus raise the surplus it offers in that market, by $\delta_{s,i}(\mathbf{z})$ as well. Hence consumers are better off than before the regulation.

4.2.3 Data cooperative

We now consider a policy that enables single-product firms to share data with each other via a data cooperative (while continuing to choose prices and qualities independently to maximize their own payoff). When all single-product firms join the cooperative, a consumer who buys single-product firm i's product now gets surplus

$$q_{s,i} - p_{s,i} + \delta_{s,i}(\mathbf{1} - \mathbf{z}) + \varepsilon_{s,i},$$

where $\delta_{s,i}(\mathbf{1} - \mathbf{z})$ is again a positive and increasing functional that captures the utility boost from data shared by the other single-product firms. We assume $\delta_{s,i}(\mathbf{0}) = 0$.

We distinguish between whether single-product firms exchange data for free, or compensate each other for it. In the latter case, the policymaker publicly proposes a trading scheme $\{t_i(\mathbf{1} - \mathbf{z})\}_{i \in \mathcal{I}}$, where $t_i(\mathbf{1} - \mathbf{z})$ is single-product firm i's *net* revenue from data trading. We assume that if all single-product firms join, the cooperative breaks even, i.e., net revenues from data trading cancel out in the aggregate. Irrespective of whether there is compensation, the timing of the extended game is as follows: each single-product firm simultaneously and publicly decides whether to join the cooperative; all firms set their prices and qualities, and then consumers make their choices; finally, where relevant, data compensation is paid.

Start with the case where single-product firms exchange their data for free. The equilibrium remains the same as before, except that now the ecosystem's net data (dis)advantage is $\delta_i(\mathbf{z}) = \delta_{s,i}(\mathbf{z}) - \delta_{s,i}(\mathbf{1} - \mathbf{z})$, and its price satisfies

$$p_{e,i} = z_i h_i(z_i) - \int z_j \left[\frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} + \frac{\partial \delta_{s,j}(\mathbf{1} - \mathbf{z})}{\partial (1 - z_i)} \right] dj. \tag{28}$$

Due to the $\frac{\partial \delta_{s,j}(\mathbf{1}-\mathbf{z})}{\partial(1-z_i)}$ term, the ecosystem prices more aggressively compared to when there is no cooperative. Intuitively, when the ecosystem cuts its prices, it now deprives the cooperative of data, reducing the surplus that its single-product rivals can offer. The equilibrium sales equation (19) now becomes

$$G_i(z_i) = \int z_j \frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} dj + \delta_{e,i}(\mathbf{z}) + \tau_i(\mathbf{z}), \tag{29}$$

where

$$\tau_i(\mathbf{z}) \equiv \int z_j \frac{\partial \delta_{s,j}(\mathbf{1} - \mathbf{z})}{\partial (1 - z_i)} dj - \delta_{s,i}(\mathbf{1} - \mathbf{z}) \tag{30}$$

captures the impact of the data cooperative on equilibrium sales in market i. It can be positive or negative: the ecosystem charges lower prices, which tends to increase z_i, but single-product firms use each other's data to offer better products, which tends to decrease z_i. Using Proposition 2, the sign of $\tau_i(\mathbf{z})$ determines whether ecosystem sales rise or fall in market i. Since each single-product firm's profit decreases in the ecosystem's sales, it also determines whether single-product firms benefit from forming a data cooperative and wish to join it. Let $\mathbf{z}^*$ and $\tilde{\mathbf{z}}^*$ be respectively the equilibrium sales before and after the data cooperative with free data exchange. Then:[34]

Proposition 4. *Suppose that, for each i, the right-hand side of (29) is strictly increasing in $\mathbf{z}$ both with and without the $\tau_i(\mathbf{z})$ term.*
(i) Single-product firms are better (worse) off from forming a data cooperative with free

[34] Similar to Proposition 2, the condition that the right-hand side of (29) is strictly increasing holds, for example, when the data effects $\delta_{e,i}(\mathbf{z})$ and $\delta_{s,i}(\mathbf{1} - \mathbf{z})$ are both linear.

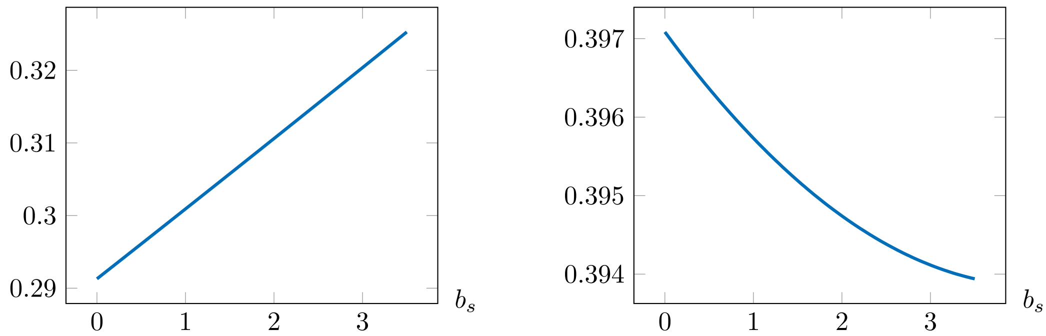
data exchange if $\tau_i(\mathbf{z}^*) < 0$ ($\tau_i(\mathbf{z}^*) > 0$) in each market. It is an equilibrium for each single-product firm to join the cooperative if and only if $\tau_i(\tilde{\mathbf{z}}^*) \leq 0$ in each market.
(ii) If all single-product firms join the cooperative, there exist compensation schemes such that consumer surplus is strictly higher relative to no data cooperative. If $\tau_i(\mathbf{z}^*) < 0$ there exist such schemes where it is also an equilibrium for each single-product firm to join the cooperative.

The intuition for why the sign of $\tau_i(\mathbf{z}^*)$ affects whether single-product firms gain or lose from being in a data cooperative was explained above. Note that to check whether it is an equilibrium for all single-product firms to join the cooperative (i.e., to check that none of them has a unilateral incentive to leave) we need to evaluate $\tau_i(\mathbf{z})$ at the equilibrium sales $\tilde{\mathbf{z}}^*$ that arise with the cooperative.

Consumers are not necessarily better off when single-product firms form a data cooperative. On the one hand, single-product firms sell more and so invest more in quality, while the ecosystem charges lower prices; on the other hand, single-product firms also charge higher prices, while the ecosystem sells less and so invests less in quality. In general either effect can dominate, as shown by Figure 5. In the left panel markets are symmetric and the single-product firms are much more efficient at quality investment than the ecosystem. Here a cooperative benefits consumers by strongly expanding single-product firms' investment. In the right panel markets are asymmetric, and in some markets the ecosystem is more efficient at quality investment than the single-product firms. Here a cooperative harms consumers by strongly reducing ecosystem investment in some of the markets.

Part (ii) of Proposition 4 then shows that if all single-product firms join the data cooperative, we can find compensation schemes for data trading such that consumers are better off than before the regulation. We follow a similar approach to the one in Section 4.2.2, by constructing compensation schemes such that equilibrium sales are the same as before the cooperative forms. Consumers in each market are then strictly better off: the ecosystem charges lower prices by (28), while single-product firms use each other's data to offer better products. If $\tau_i(\mathbf{z}^*) < 0$, we further show that it is indeed an equilibrium for all single-product firms to join the cooperative.

Finally, we note that intuitively $\tau_i(\mathbf{z}) < 0$ is more likely to hold when $\mathbf{z}$ is smaller—because the ecosystem's additional data subsidy term in (30) tends to be smaller, while single-product firms can use each other's data to offer a larger utility boost. This can be illustrated in a symmetric version of the model:



(a) Example with symmetric markets (b) Example with asymmetric markets

Figure 5: The impact of a data cooperative on consumer surplus
(F is uniform on $[-1,1]$, $\delta_{e,i}(\mathbf{z}) = b_{e,i} \int a_j z_j dj$, and $\delta_{s,i}(\mathbf{1}-\mathbf{z}) = b_{s,i} \int a_j(1-z_j)dj$. In the left panel markets are symmetric, with $\eta_e = 0.15, \eta_s = 2.6, a = 0.01$ and $b_e = 0.05$. In the right panel markets are asymmetric: 60% have the same η_e, η_s, a, b_e as in the left panel, while the remaining 40% have $\eta_e = 2.95, \eta_s = 2.8, a = 0, b_e = 3$; b_s is the same in all markets.)

Corollary 1. *Consider the case with symmetric markets, and let $\delta_s(1-z) \equiv \delta_{s,i}(\mathbf{1}-\mathbf{z})$ when $z_i = z$ for all i. If $\delta_s(\cdot)$ is log-concave and $\delta_s'(0) > 0$, there exists $\hat{z} \in (0,1)$ such that in the symmetric equilibrium $\tau_i(\mathbf{z}^*), \tau_i(\tilde{\mathbf{z}}^*) < 0$ if and only if $z^* < \hat{z}$. Moreover, $\hat{z} > \frac{1}{2}$ if $\delta_s(\cdot)$ is concave, $\hat{z} < \frac{1}{2}$ if $\delta_s(\cdot)$ is convex, and $\hat{z} = \frac{1}{2}$ if $\delta_s(\cdot)$ is linear.*

Hence, according to Proposition 4, when markets are symmetric and single-product firms exchange their data for free, a cooperative emerges and benefits its members if and only if the single-product firms are relatively large prior to the regulation.

5 A Tractable Discrete Case

As noted earlier, the general discrete case studied in Sections 2 and 3 is not sufficiently tractable; however, there is a special "symmetric double-linear" case in which we can make further progress. Suppose now that all the markets are symmetric (i.e., for each i, $F_i = F$, $\eta_{e,i} = \eta_e$, $\eta_{s,i} = \eta_s$, and $\delta_i(\mathbf{z}) = \delta(\mathbf{z})$), each market features a linear Hotelling model with $F(x) = (l+x)/(2l)$, and the data effect is linear and given by

$$\delta_i(\mathbf{z}) = w_e z_i + c \sum_{j \neq i} z_j - w_s(1-z_i), \tag{31}$$

where w_e and w_s are the within-market effect coefficients and c is the cross-market effect coefficient. For simplicity, we focus on positive data effects with $w_e, w_s, c \geq 0$.

In this special case, Assumption 1 or its equivalent condition (2) simplifies to

$$w + (n-1)c < 2l(1-\epsilon) \quad \text{with} \quad w \equiv w_e + w_s. \tag{32}$$

Note that this condition requires that the total cross-market effect does not explode as n increases. In the limit case as $n \to \infty$, it requires $(n-1)c$ to be bounded by $2l - w$, implying $c \to 0$. (This can occur when, as n increases, data from different markets become more substitutable with one another and have a smaller marginal effect on product improvement.)

For convenience, we use the following notation:

$$L \equiv 2l - w \quad \text{and} \quad \xi \equiv \frac{c}{L-(n-2)c} \in [0,1), \tag{33}$$

where L is interpreted as the effective degree of product differentiation in a single market with the standard within-market data effect, and the range of ξ is from condition (32).

We have the following result concerning equilibrium existence and uniqueness (all omitted proofs and details for this section can be found in the Online Appendix B.3):

Proposition 5. *Consider the discrete symmetric-double-linear case, and suppose that*

$$\max\left\{\frac{\eta_e}{2} + (n-1)c, \frac{\eta_s}{2(1-\xi)} - c\right\} \leq L, \tag{34}$$

and

$$\max\{\eta_e + w_e + 2(n-1)c, \eta_s + w_s + (n-1)c\xi\} < L + l. \tag{35}$$

There exists a unique interior symmetric equilibrium where $q_e = \eta_e z$, $q_s = \eta_s(1-z)$, $p_e = z[L-(n-1)c]$, $p_s = (1-z)[L-(n-1)c\xi]$, *and* $z = Z_s/(Z_s+Z_e) \in (0,1)$ *with* $Z_s = L + l - w_s - \eta_s - (n-1)c\xi$ *and* $Z_e = L + l - \eta_e - w_e - 2(n-1)c$.

Condition (34) ensures the second-order conditions for both the ecosystem and single-product firms, and condition (35) ensures a unique interior equilibrium. (Note also that (34) implies (32).) All the required conditions hold if the degree of product differentiation l is sufficiently large.

Both prices are positive since condition (32) implies $L > (n-1)c$ and $\xi < 1$. The $(n-1)c$ term in p_e captures the cross-market effect and the $(n-1)c\xi$ term in p_s captures the feedback effect. Since $\xi < 1$, in this special case the (per-sale) cross-market effect in p_e is stronger than the (per-sale) feedback effect in p_s. Recall that as $n \to \infty$, $(n-1)c$ is bounded and $c \to 0$, so $(n-1)c\xi \to 0$. Therefore, as seen in the continuum case, the feedback effect vanishes in the limit.

Policy implications. Here in the main text we focus on data regulation that reduces the cross-market data effect. We discuss the other two policies in the Online Appendix B.3. Recall that in the continuum case, the symmetric-double-linear specification satisfies the conditions in Proposition 2, so reducing c increases each single-product firm's sales $1-z$ and also their profit.

A major difference in the discrete case is that, because of the feedback effect $(n-1)c\xi$ in p_s (where ξ increases in c), data regulation that reduces c now induces single-product firms to raise their prices. This, in turn, can reduce their sales and profits. More specifically, one can check that data regulation increases $1-z$ if and only if

$$z > \frac{2\xi + (n-2)\xi^2}{2 + 2\xi + (n-2)\xi^2} \in (0,1). \tag{36}$$

Therefore, if the ecosystem's initial sales are relatively low, data regulation reduces single-product firms' sales. For the impact on their profits, note that $\pi_s = (1-z)^2[L - (n-1)c\xi - \eta_s/2]$. A difference with the continuum case is that regulation now affects π_s both directly (via the $(n-1)c\xi$ term) and indirectly (via z). If (36) holds and reducing c increases $1-z$, regulation must benefit single-product firms, just like in the continuum case. If instead (36) fails, it is possible that reducing c harms single-product firms, as illustrated in a numerical example reported in the Online Appendix B.3. However, both of the above differences disappear in the limit when $n \to \infty$, as both the right-hand side of (36) and $(n-1)c\xi$ in π_s converge to zero. Therefore, in general the insights from the continuum model carry over when the number of markets is large.[35]

The ecosystem's profit is $\Pi_e = z^2[L - (n-1)c - \eta_e/2]$. Since reducing c now can increase z (which happens when (36) fails), compared to the continuum case there is an additional channel through which data regulation can potentially benefit the ecosystem. However, just like in the continuum case, while it is possible for data regulation to benefit all firms, it is impossible for data regulation to harm all firms, as proved in the Online Appendix B.3.

Consumer surplus in each market is

$$V = q_s - p_s + w_s(1-z) + \int_0^{F^{-1}(z)} F(x)dx = (1-z)[\eta_s + w_s + (n-1)c\xi - L] + \left(z^2 - \frac{1}{4}\right) l$$

[35]More precisely, as $n \to \infty$, $(n-1)c$ is bounded under Assumption 1, so we assume it converges to a constant $C \in (0, L)$. Data regulation then reduces C. For large n, it is without loss of generality to consider $c = \frac{C}{n-1}$. Then for given parameters (other than c as it is now a function of n), there exists $\hat{n}$ such that (36) holds for $n > \hat{n}$, as z is bounded away from 0 while the right-hand side converges to 0.

where the second equality used the expressions for q_s and p_s and the uniform distribution F. Compared to the continuum case, the main difference is that, given $c\xi$ increases in c, data regulation now has a direct negative effect on V as it mitigates the feedback effect and so raises p_s. However, it is still possible for data regulation to benefit consumers. We show in the Online Appendix B.3 that in the limit case when $c \to 0$ or $n \to \infty$, data regulation benefits consumers if η_s is large enough, i.e., if single-product firms are sufficiently good at innovating, the same insight as in the continuum case.

6 Conclusion

This paper makes two contributions. First, it develops a framework to study competition between a multiproduct ecosystem and multiple single-product rivals across different markets. The framework features cross-market data usage by the ecosystem (which generates demand externalities across products), allows for competition in both prices and innovation, and accommodates rich product heterogeneity. It can also accommodate non-data spillovers across markets in more traditional industries.

Second, the paper uses this framework to evaluate several data policies. Restricting the ecosystem's cross-market data usage and mandating data sharing from the ecosystem to smaller competitors both tend to benefit single-product firms and encourage them to innovate more. However, these policies also dampen the ecosystem's incentive to collect data, reducing its investment and potentially leading to higher prices. As a result, they benefit consumers only when single-product firms are sufficiently good at innovating. Facilitating data exchange among single-product firms through a data cooperative can potentially harm those firms, as it induces more aggressive pricing by the ecosystem. It may also harm consumers by reducing the ecosystem's innovation. Nevertheless, for both the second and third policies, there exist compensation schemes for data sharing or data trading under which consumers are better off.

Our framework ould be further developed to address other interesting questions. For instance, instead of focusing on policies that regulate the use of data, we could examine policies that affect market structure. Besides breaking up ecosystems, policymakers could also allow single-product firms to merge and become larger competitors. Moreover, we have exogenously fixed the ecosystem's product range, but it would be interesting to endogenize its choice of which products to supply—for example by acquiring start-ups—in which case we could consider policies such as restricting mergers that involve big tech companies. 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 is 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.

A Appendix: Omitted Proofs

A.1 Proof of Lemma 1

We prove the monotonicity of z_i in v_i and v_j. We can rewrite (3) as $F_i^{-1}(z_i)-\delta_i(\mathbf{z}) = v_i$. For convenience, we use the notation $\delta_{ij}(\mathbf{z}) \equiv \frac{\partial \delta_i(\mathbf{z})}{\partial z_j}$, and let

$$J = \begin{bmatrix} h_1 - \delta_{11} & -\delta_{12} & \cdots & -\delta_{1n} \\ -\delta_{21} & h_2 - \delta_{22} & \cdots & -\delta_{2n} \\ \vdots & \vdots & & \vdots \\ -\delta_{n1} & -\delta_{n2} & \cdots & h_n - \delta_{nn} \end{bmatrix} \tag{37}$$

be the Jacobian of $\{F_i^{-1}(z_i) - \delta_i(\mathbf{z})\}$, where we have suppressed the dependent variable $\mathbf{z}$ and $h_i = h_i(z_i) \equiv 1/f_i(F_i^{-1}(z_i))$. Single-peakedness of f_i and Assumption 1 imply that $h_i \geq 1/f_i(0) > \sum_{j=1}^n |\delta_{ij}|$, so $h_i - \delta_{ii} \geq h_i - |\delta_{ii}| > \sum_{j\neq i} |\delta_{ij}| \geq 0$. Therefore, each diagonal entry in J is strictly positive, and J is strictly *diagonally dominant by rows*. This implies that (i) $|J| > 0$ and so its inverse J^{-1} exists, (ii) all diagonal entries of J^{-1} are strictly positive, and (iii) if $\delta_{ij} > 0$ for all $i \neq j$, every off-diagonal entry of J^{-1} is also strictly positive.

Let $z_{ij} \equiv \frac{\partial z_i}{\partial v_j}$, and define the $n \times n$ matrix $Z = [z_{ij}]$. Differentiating the system of equations $F_i^{-1}(z_i) - \delta_i(\mathbf{z}) = v_i$ with respect to $\mathbf{v} = (v_1, \cdots, v_n)$ yields $JZ = I$, and so $Z = J^{-1}$. Therefore, $z_{ij} = (J^{-1})_{ij}$, and it must be strictly positive for $i = j$ by the above property (ii). Moreover, if all $\delta_i(\mathbf{z})$'s are increasing and so $\delta_{ij} > 0$, then by the above property (iii) z_{ij} is also strictly positive for $i \neq j$.[36] □

[36] Alternatively, when all $\delta_i(\mathbf{z})$'s are increasing functions of $\mathbf{z}$, $z_i = F_i(v_i+\delta_i(\mathbf{z}))$ defines an increasing

A.2 Proof of Proposition 1

Given the derived second-order conditions, it suffices to prove the following result:

Claim 1. *If, for each $i \in [0,1]$, conditions* (20) *and* (21) *hold, then the system of equations (19) has a unique interior solution with $z_i \in (0,1)$ for each $i \in [0,1]$.*

Recall that the equilibrium sales are determined by $G_i(z_i) = \Delta_i(\mathbf{z})$. A sufficient condition for this system to have interior solutions is $G_i(0) < \Delta_i(\mathbf{z})$ and $G_i(1) > \Delta_i(\mathbf{z})$ for any i and $\mathbf{z}$. (In the right-hand side of each inequality, $\mathbf{z}$ should be conditional on $z_i = 0$ and $z_i = 1$, respectively. This, however, does not matter in our continuum setup.) Given the symmetry of f_i, they can be written as, for any i,

$$G_i(0) = \eta_{s,i} - l_i - \frac{1}{f_i(l_i)} < \min_{\mathbf{z}} \Delta_i(\mathbf{z}) \quad \text{and} \quad G_i(1) = l_i + \frac{1}{f_i(l_i)} - \eta_{e,i} > \max_{\mathbf{z}} \Delta_i(\mathbf{z}).$$

They lead to condition (20).

To prove uniqueness, let us introduce notation $\theta(\mathbf{z}) \equiv \int z_j \delta_j(\mathbf{z}) dj$. Then the equilibrium sales equation can be written as $G_i(z_i) = \frac{\partial \theta(\mathbf{z})}{\partial z_i}$. Define a potential function

$$\Gamma(\mathbf{z}) \equiv \int \left\{ \int_0^{z_i} (-G_i(t_i)) dt_i \right\} di + \theta(\mathbf{z}). \tag{38}$$

It is Fréchet differentiable, and $\frac{\partial \Gamma(\mathbf{z})}{\partial z_i} = -G_i(z_i) + \frac{\partial \theta(\mathbf{z})}{\partial z_i}$. Therefore, if $\Gamma(\mathbf{z})$ is strictly concave, the system of equilibrium sales equations has a unique solution. We claim that under condition (21), $\Gamma(\cdot)$ is indeed strictly concave. To see that, we add and subtract a quadratic term $\frac{\lambda}{2} \int z_i^2 di$ in Γ:

$$\Gamma(\mathbf{z}) = \int \left\{ \int_0^{z_i} (\lambda t_i - G_i(t_i)) dt_i \right\} di + \left(\theta(\mathbf{z}) - \frac{\lambda}{2} \int z_i^2 di \right). \tag{39}$$

The second term is concave given the definition of λ. The first term is strictly concave in $\mathbf{z}$ if $\int_0^{z_i} (\lambda t_i - G_i(t_i)) dt_i$ is strictly concave in z_i for each i. This is true as its second derivative is $\lambda - g_i(z_i) \leq \lambda - (\frac{3}{f_i(0)} - 2\bar{\eta}_i) < 0$, where the second inequality is from condition (21). To see the first inequality, notice that from (14), we have

$$g_i(z_i) = 3h_i(z_i) - 2\bar{\eta}_i + (2z_i - 1)h_i'(z_i) \geq \frac{3}{f_i(0)} - 2\bar{\eta}_i, \tag{40}$$

where the inequality follows because the symmetry and log-concavity of f_i imply that $f_i(0) \geq 1/h_i(z_i)$ and also $(2z_i - 1)h_i'(z_i) \geq 0$. Thus, Γ, as the sum of these two terms, is strictly concave. □

mapping. Given the uniqueness of the solution, standard monotone comparative statics imply that z_i increases in v_j for all j.

A.3 Proof of Proposition 2

Suppose that regulation reduces the data spillovers from $\{\delta_i(\mathbf{z})\}$ to $\{\tilde{\delta}_i(\mathbf{z})\}$, with both satisfying the conditions in Proposition 1. Let $\Delta_i(\mathbf{z})$ and $\tilde{\Delta}_i(\mathbf{z})$ be the corresponding data-driven benefits of buying the ecosystem's product i defined in the right-hand side of (19). Let $\mathbf{z}^*$ and $\tilde{\mathbf{z}}^*$ be the corresponding equilibrium sales which solve (19).

Lemma 4. *If, for each i, both $\Delta_i(\mathbf{z})$ and $\tilde{\Delta}_i(\mathbf{z})$ are increasing functions and*

$$\tilde{\Delta}_i(\mathbf{z}^*) \leq \Delta_i(\mathbf{z}^*), \tag{41}$$

then $\tilde{\mathbf{z}}^* \leq \mathbf{z}^*$. If (41) is strict in a positive measure of markets, then $\tilde{z}_i^* < z_i^*$ at least in those markets; moreover, if both $\Delta_i(\mathbf{z})$ and $\tilde{\Delta}_i(\mathbf{z})$ are further strictly increasing functions for each i, then $\tilde{z}_i^* < z_i^*$ in all the markets. (The opposite is true if $\tilde{\Delta}_i(\mathbf{z}^*) \geq \Delta_i(\mathbf{z}^*)$ for each i.)

Proof. Let $T_i(\mathbf{z}) \equiv G_i^{-1}(\Delta_i(\mathbf{z}))$. (This function is well defined given $G_i(\cdot)$ is a strictly increasing function.) Then (19) can be written as $\mathbf{z} = T(\mathbf{z})$, where $T = \{T_i\}_{i\in[0,1]}$. Similarly, define $\tilde{T}_i(\mathbf{z}) \equiv G_i^{-1}(\tilde{\Delta}_i(\mathbf{z}))$ and $\tilde{T} = \{\tilde{T}_i\}_{i\in[0,1]}$.

Note that $\mathbf{z}^* = T(\mathbf{z}^*) \geq \tilde{T}(\mathbf{z}^*)$, where the inequality is because of (41) and because $G_i^{-1}(\cdot)$ is increasing. When both Δ_i and $\tilde{\Delta}_i$ are increasing functions, both $T(\cdot)$ and $\tilde{T}(\cdot)$ are increasing mappings. Applying $\tilde{T}$ to both sides of the above inequality yields $\tilde{T}(\mathbf{z}^*) \geq \tilde{T}(\tilde{T}(\mathbf{z}^*)) \equiv \tilde{T}^2(\mathbf{z}^*)$. Successively applying this operation gives

$$\mathbf{z}^* \geq \tilde{T}(\mathbf{z}^*) \geq \tilde{T}^2(\mathbf{z}^*) \geq \cdots \geq \tilde{T}^n(\mathbf{z}^*) \geq \cdots$$

This monotone sequence must converge given the range of $\tilde{T}$ is $[0,1]^{\mathcal{I}}$, and the limit point must be $\tilde{\mathbf{z}}^*$ given $\mathbf{z} = \tilde{T}(\mathbf{z})$ has a unique fixed point $\tilde{\mathbf{z}}^*$. Therefore, $\tilde{\mathbf{z}}^* \leq \mathbf{z}^*$.[37]

Given $\tilde{\mathbf{z}}^* \leq \mathbf{z}^*$, we have

$$G_i(\tilde{z}_i^*) = \tilde{\Delta}_i(\tilde{\mathbf{z}}^*) \leq \tilde{\Delta}_i(\mathbf{z}^*) \leq \Delta_i(\mathbf{z}^*) = G_i(z_i^*), \tag{42}$$

where the first inequality is because $\tilde{\Delta}(\cdot)$ is increasing and the second one is because of (41). When (41) is strict in market i, the second inequality is strict and so $\tilde{z}_i^* < z_i^*$. Now consider market j where (41) is not strict. Given $\tilde{z}_i^* < z_i^*$ in a positive measure of

[37] This argument does not work when both Δ_i and $\tilde{\Delta}_i$ are *decreasing* functions. In that case, we can only show that $\tilde{z}_i^* \leq z_i^*$ in a positive measure of markets (but not necessarily in all the markets). To see that, suppose in contrast $\tilde{\mathbf{z}}^* > \mathbf{z}^*$. Then $\tilde{\mathbf{z}}^* > \mathbf{z}^* = T(\mathbf{z}^*) \geq \tilde{T}(\mathbf{z}^*) \geq \tilde{T}(\tilde{\mathbf{z}}^*)$, where the last inequality used that $\tilde{T}$ is now a decreasing mapping. This contradicts $\tilde{\mathbf{z}}^* = \tilde{T}(\tilde{\mathbf{z}}^*)$.

markets and $\tilde{\Delta}_j(\mathbf{z})$ is strictly increasing, the first inequality in the market-j version of (42) will be strict and so $\tilde{z}_j^* < z_j^*$. The proof is similar when $\tilde{\Delta}_i(\mathbf{z}^*) \geq \Delta_i(\mathbf{z}^*)$. □

The impact of regulation on investments, prices, and profits, has already been explained in the main text. For consumer surplus, using (25), we have

$$\frac{dV_i}{dz_i} = -\eta_{s,i} + (1+z_i)h_i(z_i) - (1-z_i)h_i'(z_i).$$

Using $\sigma_i(\cdot)$ defined in (5), one can check that this is negative if (22) holds. When this is true for all $z_i \in [\tilde{z}_i^*, z_i^*]$, we have $V_i(\tilde{z}_i^*) > V_i(z_i^*)$. □

A.4 Proof of Proposition 3

Part (i) has been explained in the text, so the proof is omitted. Consider part (ii). Start with a general compensation scheme $\{t_i(\mathbf{z})\}_{i\in\mathcal{I}}$ where no single-product firm's choices affect its payment. We look for an equilibrium where all single-product firms buy data. If all single-product firms do indeed buy data, the equilibrium of the consumer choice game is again given by a continuum of equations $z_i = F_i(v_i + \delta_i(\mathbf{z}))$, except that now $\delta_i(\mathbf{z}) = \delta_{e,i}(\mathbf{z}) - \delta_{s,i}(\mathbf{z})$. Single-product firm i's profit is $p_{s,i}(1-z_i) - C_{s,i}(q_{s,i}) - t_i(\mathbf{z})$. Since its behavior has no impact on $t_i(\mathbf{z})$, its optimal price and quality are the same as under no data sharing, i.e., $q_{s,i} = \eta_{s,i}(1-z_i)$ and $p_{s,i} = (1-z_i)h_i(z_i)$.

The ecosystem's profit is $\int[p_{e,i}z_i - C_{e,i}(q_{e,i}) + t_i(\mathbf{z})]di$. Following the same approach as before we have $q_{e,i} = \eta_{e,i}z_i$ and

$$p_{e,i} = z_i h_i(z_i) - \int \left[z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} + \frac{\partial t_j(\mathbf{z})}{\partial z_i} \right] dj. \tag{43}$$

Hence equilibrium sales are determined by a continuum of equations

$$G_i(z_i) = \int \left[z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} + \frac{\partial t_j(\mathbf{z})}{\partial z_i} \right] dj + \delta_i(\mathbf{z}). \tag{44}$$

Now consider the specific compensation scheme $t_j(\mathbf{z}) = k_j + \int z_l \delta_{s,l}(\mathbf{z}) dl$ where k_j is a market-specific fixed fee. With this scheme, a change in z_j indeed has no impact on $t_j(\mathbf{z})$, but its Fréchet derivative is well defined and we have

$$\int \frac{\partial t_j(\mathbf{z})}{\partial z_i} dj = \delta_{s,i}(\mathbf{z}) + \int z_j \frac{\partial \delta_{s,j}(\mathbf{z})}{\partial z_i} dj \tag{45}$$

because the k_j are fixed. Substituting (45) into (44) gives

$$G_i(z_i) = \int z_j \frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} dj + \delta_{e,i}(\mathbf{z}), \tag{46}$$

which coincides with (19) evaluated at $\delta_i(\mathbf{z}) = \delta_{e,i}(\mathbf{z})$. Hence equilibrium $\mathbf{z}$ is the same with or without compensated data sharing.[38] Hence compensated data sharing raises consumer surplus in any market i by $\delta_{s,i}(\mathbf{z})$ because (i) single-product firm i's surplus is $\delta_{s,i}(\mathbf{z})$ higher due to the shared data, and (ii) by substituting (45) into (43),

$$p_{e,i} = z_i h_i(z_i) - \int z_j \frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} dj - \delta_{s,i}(\mathbf{z}),$$

which is lower by $\delta_{s,i}(\mathbf{z})$ compared to the case with no regulation.

It remains to prove that under the compensation scheme (27) each single-product firm wishes to buy the data, given that all other single-product firms choose to do so. To this end, suppose single-product firm i deviates from the putative equilibrium and chooses not to buy the ecosystem's data. Consumers observe this, and so for a given profile of prices and qualities, the consumer choice game is characterized by

$$\tilde{z}_i = F_i(v_i + \delta_{e,i}(\mathbf{z})) \quad \text{and} \quad z_j = F_j(v_j + \delta_j(\mathbf{z})) \quad \text{for } j \neq i.$$

Since a change in z_i has no impact on $\delta_j(\mathbf{z})$ for any $j \neq i$, the solution to this new system comprises the $\tilde{z}_i$ defined above, and the $\{z_j\}_{j \neq i}$ that solve the system of demand equations in the putative equilibrium. Thus, in deviation market i, $q_{e,i} = \eta_{e,i}\tilde{z}_i$ and

$$p_{e,i} = \tilde{z}_i h_i(\tilde{z}_i) - \int \left[z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} + \frac{\partial t_j(\mathbf{z})}{\partial z_i} \right] dj = \tilde{z}_i h_i(\tilde{z}_i) - \int z_j \frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} dj - \delta_{s,i}(\mathbf{z}),$$

where the subsidy term in the first expression remains the same as in the putative equilibrium where all single-product firms accept, and the second equality uses (45). (Notice that in the putative equilibrium $p_{e,i}$ is the same except that in the first term $\tilde{z}_i$ is replaced by z_i.) For single-product firm i, as usual we have $q_{s,i} = \eta_{s,i}(1 - \tilde{z}_i)$ and $p_{s,i} = (1 - \tilde{z}_i)h_i(\tilde{z}_i)$. Hence $\tilde{z}_i$ in the deviation equilibrium solves

$$G_i(\tilde{z}_i) = \int z_j \frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} dj + \delta_{s,i}(\mathbf{z}) + \delta_{e,i}(\mathbf{z}).$$

[38]With this compensation scheme, the ecosystem's second-order condition remains unchanged provided that we write $\theta(\mathbf{z}) = \int z_i[\delta_{e,i}(\mathbf{z}) + \delta_{s,i}(\mathbf{z})]di$ in the definition of λ in footnote 25. The conditions for (46) to have a unique interior solution are then the usual ones.

Comparing with (46), we observe that $G_i(\tilde{z}_i) - G_i(z_i) = \delta_{s,i}(\mathbf{z})$. Since G_i is a strictly increasing function, this implies that $\tilde{z}_i > z_i$, i.e., turning down the offer leads to less sales (and so lower profit from the product market) for single-product firm i. Single-product firm i therefore accepts the compensation scheme if it satisfies $t_i(\mathbf{z}) < \pi_{s,i}(z_i) - \pi_{s,i}(\tilde{z}_i)$, where $\pi_{s,i}(\cdot)$ is single-product firm i's profit defined in (23) and it is a strictly decreasing function. Given the right-hand side is strictly positive, this inequality is satisfied for a range of strictly positive $t_i(\mathbf{z})$, i.e., provided k_i is sufficiently close to $-\int z_j \delta_{s,j}(\mathbf{z}) dj$ where $\mathbf{z}$ is equilibrium sales absent data sharing (and hence also equilibrium sales in the putative equilibrium where all single-product firms accept the offer). □

A.5 Proof of Proposition 4

Start with part (i). First, suppose all single-product firms join the cooperative, and check when each one is better off compared to no cooperative. The equilibrium analysis in Section 4.1 carries over except that now $\delta_i(\mathbf{z}) = \delta_{e,i}(\mathbf{z}) - \delta_{s,i}(\mathbf{1} - \mathbf{z})$, so $p_{e,i}$ satisfies (28) and equilibrium sales $\tilde{\mathbf{z}}^*$ satisfy (29). Hence, letting $\tilde{\Delta}_i(\cdot)$ be the right-hand side of (29), we have that $\tilde{\Delta}_i(\mathbf{z}) < \Delta_i(\mathbf{z})$ if and only if $\tau_i(\mathbf{z}) < 0$. Therefore, Lemma 4 on page 38 implies that if $\tau_i(\mathbf{z}^*) < 0$ for all i, then $\tilde{z}_i^* < z_i^*$ holds for all i. Since $\pi_{s,i}$ is strictly decreasing in z_i, each single-product firm earns strictly more profit than before the cooperative. (The opposite result when $\tau_i(\mathbf{z}^*) > 0$ for all i is proved in the same way.)

Second, consider when it is an equilibrium for the cooperative to exist. Specifically, suppose all single-product firms have joined, and consider the incentive of the one in market i to deviate and leave. We have just seen that if this firm stays, equilibrium sales are $\tilde{\mathbf{z}}^*$ that solve (29), and single-product firm i's sales are thus $1 - \tilde{z}_i^*$. Now suppose this firm leaves the data cooperative. The ecosystem's price in market i changes to

$$p_{e,i} = z_i h_i(z_i) - \int z_j \frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} dj,$$

the same as absent the data cooperative. This is simply because $1 - z_i$ no longer appears in $\delta_{s,j}(\mathbf{1} - \mathbf{z})$ for $j \neq i$ and so the ecosystem's price in market i has no impact on the offerings of single-product firms in other markets via the data-spillover effect. (All other price and quality expressions remain unchanged.) Moreover, single-product firm i now has no access to data from other single-product firms, so its utility boost $\delta_{s,i}(\mathbf{1} - \mathbf{z})$ vanishes. As a result, the equilibrium sales now become $\mathbf{z} = \{\tilde{\mathbf{z}}_{-i}^*, \tilde{z}_i\}$ where $\tilde{z}_i$ solves

$$G_i(\tilde{z}_i) = \int z_j \frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} dj + \delta_{e,i}(\mathbf{z}).$$

Because in this continuum case the right-hand side remains unchanged when $\mathbf{z}$ is replaced by $\tilde{\mathbf{z}}^* = \{\tilde{\mathbf{z}}^*_{-i}, \tilde{z}^*_i\}$, we have

$$G_i(\tilde{z}_i) = \int \tilde{z}^*_j \frac{\partial \delta_{e,j}(\tilde{\mathbf{z}}^*)}{\partial z_i} dj + \delta_{e,i}(\tilde{\mathbf{z}}^*).$$

Since $G_i(\cdot)$ is strictly increasing, from (29) we have that $\tilde{z}_i \geq \tilde{z}^*_i$ (i.e., single-product firm i sells less and earns a lower profit under the deviation) if and only if $\tau_i(\tilde{\mathbf{z}}^*) \leq 0$.

Next, consider part (ii) of the proposition. Consider a trading mechanism under which single-product firm i's *net* revenue (which can be negative) from data trading is

$$t_i(z_i) = k_i + (1 - z_i)\tau_i(\mathbf{z}^*)$$

where k_i is a market-specific fixed revenue and $\mathbf{z}^*$ is the equilibrium sales absent data cooperative (so $\tau_i(\mathbf{z}^*)$ is treated as a constant). Suppose first that all single-product firms join the cooperative. Single-product firm i's profit is now $p_{s,i}(1 - z_i) - C_{s,i}(q_{s,i}) + t_i(z_i)$, so its optimal quality is $q_{s,i} = (1 - z_i)\eta_{s,i}$ as usual, but its optimal price is now[39]

$$p_{s,i} = (1 - z_i)h_i(z_i) - \tau_i(\mathbf{z}^*). \tag{47}$$

Note that monetary transfers between single-product firms do not directly affect the ecosystem's incentives. Hence, given single-product firms' choices, the ecosystem's optimal behavior is the same as in part (i) of this proof, i.e., $q_{e,i} = z_i\eta_{e,i}$ and $p_{e,i}$ is given by (28). Therefore equilibrium sales are determined by

$$G_i(z_i) = \int z_j \left[\frac{\partial \delta_{e,j}(\mathbf{z})}{\partial z_i} + \frac{\partial \delta_{s,j}(\mathbf{1} - \mathbf{z})}{\partial (1 - z_i)} \right] dj - \tau_i(\mathbf{z}^*) + \delta_{e,i}(\mathbf{z}) - \delta_{s,i}(\mathbf{1} - \mathbf{z}). \tag{48}$$

Given the definition of $\tau_i(\mathbf{z}^*)$, the right-hand side of this equation simplifies to the right-hand side of (19) from when there is no regulation—and so $\tilde{\mathbf{z}}^* = \mathbf{z}^*$. That is, equilibrium sales are the same as when there is no data regulation. As explained in the main text, the data cooperative must then raise consumer surplus. One way to make the cooperative break even is to set $k_i = -(1 - z^*_i)\tau_i(\mathbf{z}^*)$, such that each single-product firm breaks even in the data market.

Finally, we argue that when $\tau_i(\mathbf{z}^*) \leq 0$ in each market, it is an equilibrium that all single-product firms join the cooperative. Note first that if $\tau_i(\mathbf{z}^*) \leq 0$ in each market, in the hypothetical equilibrium with all single-product firms joining the cooperative, they are better off compared to when there is no data cooperative. This is because each has

[39] Since the proposed compensation scheme is linear in z_i, the second-order condition (18) applies.

the same sales but from (47) charges a higher price. Now suppose single-product firm i unilaterally deviates and leaves the data cooperative. In that market, $p_{e,i}$ will take the same form as when there was no data cooperative (because the term $\int z_j \frac{\partial \delta_{s,j}(\mathbf{1}-\mathbf{z})}{\partial(1-z_i)} dj$ disappears), as will $p_{s,i}$ (because the term $\tau_i(\mathbf{z}^*)$ disappears), and therefore z_i in the deviation market will be determined by (19) (because the term $\delta_{s,i}(\mathbf{1}-\mathbf{z})$ disappears as well). This implies that single-product firm i has the same sales and charges the same price, and thus earns the same profit, as when there is no data cooperative. Therefore, the deviation is unprofitable. This completes the proof. □

A.6 Proof of Corollary 1

Note that, with symmetric markets and $z_i = z$ for all i, we have

$$\tau_i(\mathbf{z}) = \int z_j \frac{\partial \delta_{s,j}(\mathbf{1}-\mathbf{z})}{\partial(1-z_i)} dj - \delta_{s,i}(\mathbf{1}-\mathbf{z}) = z\delta_s'(1-z) - \delta_s(1-z).$$

It is strictly negative at $z = 0$, strictly positive at $z = 1$ given $\delta_s'(0) > 0 = \delta_s(0)$, and single-crossing given $\delta_s'(\cdot) > 0$ and that log-concavity of $\delta_s(\cdot)$ implies that $\frac{\delta_s'(1-z)}{\delta_s(1-z)}$ increases in z. Therefore, when $z_i^* = z^*$ for all i, $\tau_i(\mathbf{z}^*) < 0$ if and only if z^* is less a threshold $\hat{z}$. When $\tau_i(\mathbf{z}^*) < 0$, we must have $\tilde{z}^* < z^*$ according to Proposition 2, and so the single-crossing property of $\tau_i(\mathbf{z})$ also implies $\tau_i(\tilde{\mathbf{z}}^*) < 0$.

Finally, if $\delta_s(.)$ is concave then $\delta_s(1-z) = \int_0^{1-z} \delta_s'(x)dx > (1-z)\delta_s'(1-z)$, so $\tau_i(\mathbf{z})$ is strictly less than $(2z-1)\delta_s'(1-z)$, which is negative for all $z \leq 1/2$, implying that $\hat{z} > 1/2$. The proof when $\delta_s(.)$ is convex or linear is similar and so omitted. □

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.

Bhargava, H., A. Dubus, D. Ronayne, and S. Shekhar (2026): "The Strategic Value of Data Sharing in Interdependent Markets," *Management Science*, 72(2), 1472–1488.

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.

Farrell, J., and P. Klemperer (2007): "Chapter 31 Coordination and Lock-In: Competition with Switching Costs and Network Effects," vol. 3 of *Handbook of Industrial Organization*, pp. 1967–2072. Elsevier.

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 (2025): "Network Centrality and Device Ecosystems," mimeo.

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 (2025): "Regulating Digital Platform Ecosystems through Data-Sharing and Data-Siloing: Consequences for Innovation and Welfare," *MIS Quarterly*, 49(1), 123–154.

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

Peitz, M., and S. Sato (2025): "Asymmetric Platform Oligopoly," *RAND Journal of Economics forthcoming*.

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.

B Online Appendix [Not For Publication]

B.1 Omitted Details in the Discrete Problem

B.1.1 The Lagrangian multipliers in the single-product firm problem

To pin down the Lagrangian multipliers $\{\mu^i_{j\neq i}\}$ in single-product firm i's problem, we use the first-order conditions with respect to $\{z_{j\neq i}\}$:

$$\mu^i_j h_j(z_j) = \sum_{l=1}^{n} \mu^i_l \frac{\partial \delta_l(\mathbf{z})}{\partial z_j}. \tag{49}$$

Together with $\mu^i_i = 1 - z_i$, for any given $\mathbf{z}$ this forms a system of linear equations in $\{\mu^i_j\}$ (including $j = i$) as

$$H_i \cdot (\mu^i_1, \ldots, \mu^i_n)^T = [0, \cdots, 1 - z_i, \cdots, 0]^T,$$

where

$$H_i = \begin{bmatrix} h_1 - \delta_{11} & -\delta_{21} & \cdots & -\delta_{i1} & \cdots & -\delta_{n1} \\ \vdots & \vdots & & \vdots & & \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \\ -\delta_{1n} & -\delta_{2n} & \cdots & -\delta_{in} & \cdots & h_n - \delta_{nn} \end{bmatrix}$$

where the i_{th} row is $(0, \cdots, 1, \cdots, 0)$ and $\delta_{lj} = \frac{\partial \delta_l(\mathbf{z})}{\partial z_j}$.

Note that if we remove the i_{th} row and column in H_i, the remaining submatrix is strictly diagonally dominant by columns and its diagonal entries are strictly positive under Assumption 1. Therefore, $|H_i| > 0$ and so H_i is invertible.

It is then immediate that

$$\mu^i_j = (1 - z_i) \left(H_i^{-1}\right)_{ji} \tag{50}$$

where $\left(H_i^{-1}\right)_{ji}$ is the ji_{th} entry of H_i^{-1}. In general, $\left(H_i^{-1}\right)_{ji}$ is a function of $\mathbf{z}$ and does not have a simple expression, except in some special cases:

(a) *Two product markets.* When there are only two markets i and j, one can solve from (49) that

$$\mu^i_j = (1 - z_i) \frac{\frac{\partial \delta_i(\mathbf{z})}{\partial z_j}}{h_j(z_j) - \frac{\partial \delta_j(\mathbf{z})}{\partial z_j}}. \tag{51}$$

However, even with this expression for μ_j^i, in general it is still hard to make progress with single-product firms' second-order conditions and equilibrium existence and uniqueness.

(b) *Symmetric product markets.* Suppose all product markets are symmetric. In a symmetric equilibrium with $z_i = z$ for all i, $\mu_j^i = \mu_l^i$ for $j, l \neq i$ and (49) simplifies to

$$\mu_j^i h(z) = (1-z)\frac{\partial \delta_i(\mathbf{z})}{\partial z_j} + \mu_j^i \frac{\partial \delta_j(\mathbf{z})}{\partial z_j} + (n-2)\mu_j^i \frac{\partial \delta_l(\mathbf{z})}{\partial z_j},$$

from which we solve

$$\mu_j^i = (1-z)\frac{\frac{\partial \delta_i(\mathbf{z})}{\partial z_j}}{h(z) - \frac{\partial \delta_j(\mathbf{z})}{\partial z_j} - (n-2)\frac{\partial \delta_i(\mathbf{z})}{\partial z_j}} \tag{52}$$

where we have used $\frac{\partial \delta_i(\mathbf{z})}{\partial z_i} = \frac{\partial \delta_j(\mathbf{z})}{\partial z_j}$ and $\frac{\partial \delta_l(\mathbf{z})}{\partial z_j} = \frac{\partial \delta_i(\mathbf{z})}{\partial z_j}$ due to symmetry.

Assumption 1 implies that the denominator in both (51) and (52) is positive, so μ_j^i has the same sign as $\frac{\partial \delta_i(\mathbf{z})}{\partial z_j}$, i.e., it is positive (negative) if cross-market data effect is positive (negative). In fact, in general all μ_j^i's are positive if all $\delta_i(\mathbf{z})$'s are increasing functions.

Claim 2. *$\mu_j^i > 0$ if all $\delta_i(\mathbf{z})$'s are strictly increasing functions.*

Proof. Without loss of generality, let us focus on $i = 1$. Then

$$H_1 = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{c} & D \end{bmatrix}$$

is a block matrix, where $\mathbf{0}^T$ is a row vector of zeros, $\mathbf{c}$ is a column vector, and D is a strictly diagonally dominant (by columns) matrix. Its inverse is

$$H_1^{-1} = \begin{bmatrix} 1 & \mathbf{0}^T \\ -D^{-1}\mathbf{c} & D^{-1} \end{bmatrix}.$$

When all $\delta_i(\mathbf{z})$'s are strictly increasing, all the off-diagonal entries in D are strictly negative, and so all the entries of D^{-1} are strictly positive. Given $\mathbf{c}$ is strictly negative everywhere, we must have $-D^{-1}\mathbf{c}$ (so all $\left(H_1^{-1}\right)_{j1}$ for $j \geq 2$) are strictly positive everywhere.[40] $\square$

[40] Unfortunately, when all $\delta_i(\mathbf{z})$'s are strictly decreasing functions, we do not have $\mu_j^i < 0$. This is because when the off-diagonal entries in D are positive, it is not always the case that D^{-1} has negative entries everywhere.

B.1.2 The second-order condition of single-product firms

To investigate a single-product firm's second-order condition, it turns out to be more convenient to use the direct approach instead of the Lagrangian approach. Single-product firm i's profit is

$$\pi_{s,i} = p_{s,i}(1 - z_i) - \frac{q_{s,i}^2}{2\eta_{s,i}}.$$

It is too restrictive to require $\pi_{s,i}$ to be globally concave in $(p_{s,i}, q_{s,i})$. For example, for a fixed $q_{s,i}$, $\pi_{s,i}$ is often not globally concave in price as we know from the Hotelling model with a general log-concave distribution. In the following, we look for conditions for $\pi_{s,i}$ to be single-peaked in $(p_{s,i}, q_{s,i})$.

Recall that the system of demand equations in (3) implicitly defines $z_i(v_1, \cdots, v_n)$, and we also know that $\frac{\partial z_i}{\partial v_i} > 0$ from Lemma 1.

Claim 3. *$\pi_{s,i}$ is single-peaked in $(p_{s,i}, q_{s,i})$ if $1 - z_i$ is log-concave in v_i and $\eta_{s,i}\frac{\partial z_i}{\partial v_i} \leq 1$.*

Proof. For convenience, we use notation $z_i' \equiv \frac{\partial z_i}{\partial v_i}$ and $z_i'' \equiv \frac{\partial^2 z_i}{\partial v_i^2}$ in this proof. Single-product firm i's first-order conditions are:

$$\frac{\partial \pi_{s,i}}{\partial p_{s,i}} = 1 - z_i - p_{s,i}z_i' = 0 \quad \text{and} \quad \frac{\partial \pi_{s,i}}{\partial q_{s,i}} = p_{s,i}z_i' - \frac{q_{s,i}}{\eta_{s,i}} = 0,$$

where have used the fact that $\frac{\partial z_i}{\partial p_{s,i}} = -\frac{\partial z_i}{\partial q_{s,i}} = z_i'$. The second-order derivatives are

$$\frac{\partial^2 \pi_{s,i}}{\partial p_{s,i}^2} = -2z_i' - p_{s,i}z_i'', \ \frac{\partial^2 \pi_{s,i}}{\partial q_{s,i}^2} = -\frac{1}{\eta_{s,i}} - p_{s,i}z_i'', \ \frac{\partial^2 \pi_{s,i}}{\partial q_{s,i}\partial p_{s,i}} = z_i' + p_{s,i}z_i''.$$

Given $\pi_{s,i}$ is a continuous function, it must be single-peaked if it is locally concave at *any* critical point, i.e., at each $(p_{s,i}, q_{s,i})$ that solves the first-order conditions.

Using $p_{s,i} = (1 - z_i)/z_i'$ from the first first-order condition, we have

$$\frac{\partial^2 \pi_{s,i}}{\partial p_{s,i}^2} \leq 0 \iff 2(z_i')^2 + (1 - z_i)z_i'' \geq 0.$$

If $1 - z_i$ is log-concave in v_i, we have $(z_i')^2 + (1 - z_i)z_i'' \geq 0$, so this condition must hold.

One can also check that

$$\frac{\partial^2 \pi_{s,i}}{\partial p_{s,i}^2}\frac{\partial^2 \pi_{s,i}}{\partial q_{s,i}^2} - \left(\frac{\partial^2 \pi_{s,i}}{\partial q_{s,i}\partial p_{s,i}}\right)^2 \geq 0 \iff \eta_{s,i}(z_i')^2 \leq 2z_i' + p_{s,i}z_i''.$$

At $p_{s,i} = (1 - z_i)/z_i'$, this is equivalent to

$$\eta_{s,i}(z_i')^2 \leq \frac{2(z_i')^2 + (1 - z_i)z_i''}{z_i'}.$$

Since $(z_i')^2 + (1 - z_i)z_i'' \geq 0$ under the log-concavity of $1 - z_i$, a sufficient condition for this inequality is $\eta_{s,i} z_i' \leq 1$.

Finally, we show that the second condition above implies the third required condition

$$\frac{\partial^2 \pi_{s,i}}{\partial q_{s,i}^2} \leq 0 \iff \frac{1}{\eta_{s,i}} + p_{s,i} z_i'' \geq 0.$$

Note that we have $p_{s,i} z_i'' \geq \eta_{s,i}(z_i')^2 - 2z_i'$ from the second condition above. Then the third condition holds if

$$\frac{1}{\eta_{s,i}} + \eta_{s,i}(z_i')^2 \geq 2z_i',$$

which must be true by the AM-GM inequality. □

It appears hard to derive more primitive conditions for the single-peakedness of $\pi_{s,i}$ because z_i'' can be rather complicated. One exception is when we have linear data spillovers and linear Hotelling product differentiation in all the markets, in which case z_i' is a constant and so $z_i'' = 0$. Then, from the above proof, we can immediately see that a simple sufficient condition for $\pi_{s,i}$ to be single-peaked (or to be concave) is $\eta_{s,i} z_i' \leq 2$.

B.1.3 The limit of the discrete case as $n \to \infty$

Here we show that, under a certain condition, the "feedback effect" in the discrete case vanishes as $n \to \infty$. As a result, the limit of the discrete case is qualitatively the same as the continuum case we study in Section 4.

For a given n, let

$$c_n \equiv \max_{j \neq i, \mathbf{z}} |\delta_{ij}(\mathbf{z})|$$

denote the maximum cross-market data effect, where $\delta_{ij} = \frac{\partial \delta_i(\mathbf{z})}{\partial z_j}$. Then we have the following result:

Claim 4. *The feedback effect $\sum_{j \neq i} \mu_j^i \delta_{ji}(\mathbf{z})$ in $p_{s,i}$ in (11) vanishes as $n \to \infty$ if*

$$\exists\, \hat{n} \text{ such that for any } n > \hat{n},\ (n-1)c_n < \min_{\mathbf{z}, 1 \leq i \leq n} \left\{ \frac{1-\epsilon}{f_i(0)} - |\delta_{ii}(\mathbf{z})| \right\}. \tag{53}$$

When markets are symmetric and data spillovers are linear (i.e., $\delta_i(\mathbf{z}) = \int c_{ij} z_j dj$ with $c_{ij} = c$ for $j \neq i$), then the required condition degenerates to Assumption 1 or equivalently (2). In general, however, this condition is stronger than (2).

Proof. For a given n and $\mathbf{z}$, without loss of generality, let $\left|\mu^i_{j_n}\right| \geq |\mu^i_l|$ for any $l \neq i$. Then (12) implies that

$$\left|\mu^i_{j_n}\right| h_{j_n} \leq \left|\mu^i_{j_n}\right| |\delta_{j_n j_n}| + (1 - z_i)c_n + (n-2)\left|\mu^i_{j_n}\right| c_n.$$

where we have used $\mu^i_i = 1 - z_i$ and that $x = y + z$ implies $|x| \leq |y| + |z|$. Then, for a sufficiently large n, we have

$$\left|\mu^i_{j_n}\right| \leq \frac{(1 - z_i)c_n}{h_{j_n} - |\delta_{j_n j_n}| - (n-2)c_n}$$

where the inequality holds because the denominator on the right-hand side is positive (and bounded away from zero) under condition (53). Then the feedback effect in $p_{s,i}$ satisfies

$$\left|\sum_{j \neq i} \mu^i_j \delta_{ji}\right| \leq (n-1)\left|\mu^i_{j_n}\right| c_n \leq \frac{(n-1)(1-z_i)c_n^2}{h_{j_n} - |\delta_{j_n j_n}| - (n-2)c_n}$$

which must converge to 0 as $(n-1)c_n$ is bounded and $c_n \to 0$ under condition (53). □

Note that in the limit $p_{s,i}$ can still have the within-market-effect term $(1 - z_i)\delta_{ii}(\mathbf{z})$, depending on whether $\max_{\mathbf{z}} |\delta_{ii}(\mathbf{z})| \to 0$. If the within-market effect also vanishes in the limit, the situation becomes exactly the same as our continuum case. Otherwise, we can add the within-market effect back to the continuum case, and as shown in Section B.2.5, that modification does not qualitatively change the main results there.

B.2 Omitted Details in the Continuum Case

B.2.1 Fréchet Differentiability and Partial Derivative

We briefly recap the standard textbook treatment of Fréchet derivatives. Let $\mathcal{Z}$ be a normed vector space equipped with norm $\|\cdot\|$. Then $\delta(\mathbf{z}) : \mathcal{Z} \to \mathbb{R}$ is *Fréchet differentiable* at $\mathbf{z} \in \mathcal{Z}$, if, for any $\mathbf{h} \in \mathcal{Z}$, there exists a linear transformation $T_{\mathbf{z}}(\mathbf{h}) : \mathcal{Z} \to \mathbb{R}$ such that

$$\lim_{\|\mathbf{h}\| \to 0} \frac{|\ \delta(\mathbf{z} + \mathbf{h}) - \delta(\mathbf{z}) - T_{\mathbf{z}}(\mathbf{h})\ |}{\|\ \mathbf{h}\ \|} = 0.$$

$T_{\mathbf{z}}(\mathbf{h})$ is said to be the *Fréchet derivative* at $\mathbf{z}$, and it is unique. When the norm $\|\cdot\|$ is L^2, $\mathcal{Z}$ is a Hilbert space, and the Riesz Representation Theorem implies that there exists a unique function $t_{\mathbf{z}}(i) \in L^2([0,1])$ such that

$$T_{\mathbf{z}}(\mathbf{h}) = \int_0^1 t_{\mathbf{z}}(i)h(i)di.$$

This $t_{\mathbf{z}}(i)$ is our partial derivative $\frac{\partial \delta(\mathbf{z})}{\partial z_i}$. Notice that if $h(i) = 0$ except on a zero measure of i, then the Fréchet derivative $T_{\mathbf{z}}(\mathbf{h}) = 0$ even if $t_{\mathbf{z}}(i) \neq 0$ everywhere.

B.2.2 Details of the Firm Problem in the Continuum Case

For given prices and qualities, we can solve the consumer choice game by the same approach as used in the discrete case. Here we focus on the firms' problems.

Single-product firm problem. Given other firms' choices of price and quality investment, single-product firm i's problem is to

$$\max_{p_{s,i}, q_{s,i}} \; p_{s,i}(1 - z_i) - C_{s,i}(q_{s,i})$$

where $z_i = F_i(v_i + \delta_i(\mathbf{z}))$. Given that adjusting $(p_{s,i}, q_{s,i})$ does not affect $\delta_i(\mathbf{z})$ in our continuum framework as explained in footnote 24, it is straightforward to derive the first-order conditions. However, to deal with the second-order condition, it is more convenient to use the quantity approach: using $F_i^{-1}(z_i) = v_{e,i} - [q_{s,i} - p_{s,i}] + \delta_i(\mathbf{z})$ where $v_{e,i} = q_{e,i} - p_{e,i}$, we can rewrite single-product firm i's problem as

$$\max_{z_i, q_{s,i}} \; \underbrace{[F_i^{-1}(z_i) - v_{e,i} + q_{s,i} - \delta_i(\mathbf{z})]}_{p_{s,i}}(1 - z_i) - C_{s,i}(q_{s,i}).$$

Since we can treat $\delta_i(\mathbf{z})$ as a constant, the first-order conditions with respect to z_i and $q_{s,i}$ yield respectively $p_{s,i} = (1 - z_i)h_i(z_i)$ and $q_{s,i} = \eta_{s,i}(1 - z_i)$.

To investigate single-product firm i's second-order condition, notice first that its problem is strictly concave in $q_{s,i}$ for any given z_i. Using the optimal solution $q_{s,i} = \eta_{s,i}(1 - z_i)$, we can write its profit as a function of z_i:

$$[F_i^{-1}(z_i) - v_{e,i} - \delta_i(\mathbf{z})](1 - z_i) + \frac{\eta_{s,i}}{2}(1 - z_i)^2.$$

Given we can treat $\delta_i(\mathbf{z})$ as a constant, this is concave in z_i if

$$[F_i^{-1}(z_i)(1 - z_i)]'' + \eta_{s,i} \leq 0.$$

Notice that

$$-[F_i^{-1}(z_i)(1 - z_i)]'' = 2h_i(z_i) - h_i'(z_i)(1 - z_i) = h_i(z_i)[2 - \sigma_i(z_i)].$$

Therefore, if

$$\eta_{s,i} \leq \min_{z_i \in [0,1]} h_i(z_i)[2 - \sigma_i(z_i)],$$

the first-order conditions are sufficient for determining single-product firm i's optimal solution.

Ecosystem problem. Using the quantity approach as in the discrete case, we can rewrite the ecosystem's problem as

$$\max_{\mathbf{z},\mathbf{q}_e} \int \{ \underbrace{\left[q_{e,i} - F_i^{-1}(z_i) - v_{s,i} + \delta_i(\mathbf{z})\right]}_{p_{e,i}} z_i - C_{e,i}(q_{e,i}) \} di.$$

We can do point-wise optimization with respect to $\{q_{e,i}\}$ since this dimensional is separable across markets. The first-order condition with respect to $q_{e,i}$ yields $q_{e,i} = \eta_{e,i} z_i$.

The optimization with respect to $\mathbf{z}$ is less straightforward: since a change of z_i alone does not affect any $\delta_j(\mathbf{z})$, it is not meaningful to do point-wise optimization here. However, by the standard approach of calculus of variations detailed below, we can derive the first-order conditions with respect to $\mathbf{z}$:[41] for each $i \in \mathcal{I}$,

$$p_{e,i} = z_i h_i(z_i) - \int z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj. \tag{54}$$

By using the λ defined in footnote 25, we can derive the same second-order condition as in the discrete case by using the same approach.

To demonstrate the approach of calculus of variations, let us rewrite the objective function with respect to $\mathbf{z}$ as

$$\int [\phi_i(z_i) + z_i \delta_i(\mathbf{z})] di \quad \text{with} \quad \phi_i(z_i) = [q_{e,i} - F_i^{-1}(z_i) - v_{s,i}] z_i.$$

(We have ignored the $C_{e,i}(q_{e,i})$ term as it is independent of z_i.) Consider a variation $\mathbf{z} + \mathbf{h}$. Then the first-order variation of the objective function is

$$\int \left[[\phi_i'(z_i) + \delta_i(\mathbf{z})] h_i + z_i T_{\mathbf{z}}^i(\mathbf{h}) \right] di,$$

where $T_{\mathbf{z}}^i(\mathbf{h})$ is the Fréchet derivative of $\delta_i(\mathbf{z})$. With the L^2 norm, we have

$$T_{\mathbf{z}}^i(\mathbf{h}) = \int \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} h_j dj,$$

and so

$$\int z_i T_{\mathbf{z}}^i(\mathbf{h}) di = \int z_i \int \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} h_j dj di = \int \int z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj h_i di.$$

Substituting this into the variation of the objective function yields

$$\int \left[\phi_i'(z_i) + \delta_i(\mathbf{z}) + \int z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj \right] h_i di.$$

[41] By equipping the space of $\mathbf{z}$ with the weak topology, it is standard to show that the optimization problem with respect to $\mathbf{z}$ must have a solution.

This is equal to zero for any $\mathbf{h}$ if and only if

$$\phi_i'(z_i) + \delta_i(\mathbf{z}) + \int z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj = 0$$

for any i. Using $\phi_i'(z_i) = q_{e,i} - F_i^{-1}(z_i) - v_{s,i} - z_i h_i(z_i)$, we obtain

$$\underbrace{q_{e,i} - F_i^{-1}(z_i) - v_{s,i} + \delta_i(\mathbf{z})}_{p_{e,i}} - z_i h_i(z_i) + \int z_j \frac{\partial \delta_j(\mathbf{z})}{\partial z_i} dj = 0,$$

and so (54) follows.

B.2.3 Cross-Market Comparison in the Continuum Case

Here we show how, in the continuum case, prices and qualities can be compared across markets with heterogeneous data effects. Specifically, consider two markets j and k that are identical except for the ecosystem's relative data advantages in those markets $\delta_j(\mathbf{z})$ and $\delta_k(\mathbf{z})$. Note that because there is a continuum of markets, we should treat $\mathbf{z}$ as fixed when comparing the two markets. Since markets j and k differ only in their data effects, we have $G_j(z) = G_k(z)$ for all $z \in [0, 1]$. It then follows from (19) that $z_j \geq z_k$ if and only if $\Delta_j(\mathbf{z}) \geq \Delta_k(\mathbf{z})$, or equivalently

$$\int z_i \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} di + \delta_j(\mathbf{z}) \geq \int z_i \frac{\partial \delta_i(\mathbf{z})}{\partial z_k} di + \delta_k(\mathbf{z}). \tag{55}$$

Loosely speaking, on each side of the inequality, the first term captures how good a product is at *generating* data, while the second term captures how good a product is at *using* data.[42] The ecosystem therefore tends to sell more in markets that better use data—because they have a larger data-driven boost to their demand, and in markets that better generate data—because the extra data brings benefits in other markets.

Recall that investments are proportional to demand, while a single-product firm's price is increasing in its demand. Hence, when (55) holds, the ecosystem invests more in market j than in market k, while single-product firm j invests less and also charges less than single-product firm k.

The comparison of the ecosystem's price across markets j and k is more interesting. To demonstrate, consider two cases.

[42]In the double-linear setting a_i captures how good product i is at generating data, and b_i captures how good it is at using data. Note, however, that for general $\{\delta_i(\mathbf{z})\}$, whether a product is good at using or generating data can depend on $\mathbf{z}$ and hence be endogenous.

First, suppose the markets are equally good at generating data (so the first term on each side of (55) is the same) but market j is better at using data (so $\delta_j(\mathbf{z}) > \delta_k(\mathbf{z})$). Using equation (17) we see that $p_{e,j} > p_{e,k}$ because the data subsidy in each market is the same, and because $zh(z)$ strictly increases in z given log-concavity. Hence, as one might expect, the ecosystem charges more in the market that is better at using data.

Second, suppose the markets are equally good at using data (so $\delta_j(\mathbf{z}) = \delta_k(\mathbf{z})$) but market j is better at generating data (so the first term on the left-hand side of (55) is strictly larger than the first term on the right-hand side). The comparison between $p_{e,j}$ and $p_{e,k}$ is now more subtle. On the one hand, the ecosystem sells more in market j, which is a force towards charging more in that market. On the other hand, the ecosystem offers a larger data subsidy in market j. In general either effect can dominate, and so the ecosystem may charge more in the market that generates more data. To illustrate this last point, suppose that $\int z_i \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} di = \int z_i \frac{\partial \delta_i(\mathbf{z})}{\partial z_k} di + \epsilon$ for $\epsilon \approx 0$ and positive. Using a Taylor approximation, one can check that $p_{e,j} > p_{e,k}$ if and only if

$$\eta_e + \eta_s > h(z)[2 - \sigma(z)]$$

where z denotes output in the two markets at $\epsilon = 0$. (Since all non-data terms are the same in both markets, we drop market-specific subscripts on η_e, η_s, h and σ.) Hence the ecosystem charges more in market j provided it and the single-product firms are relatively good at innovating. Intuitively, the ecosystem wishes to sell more in market j, and so optimally offers higher net surplus in that market. When η_e is high, it is more profitable to do this through higher investment, rather than via a lower price. Moreover, when η_s is relatively high, the single-product firm in market j reduces its investment—and hence the net utility it offers consumers—by more in anticipation of the reduction in its demand. This makes it easier for the ecosystem to increase sales in market j through investment alone, allowing it to charge a higher price in that market.

B.2.4 Examples of non-linear data effects

Here we report two examples of non-linear data effects to further discuss the conditions required in Proposition 2. We focus on the continuum case with $\delta_i(\mathbf{z}) = \delta_{e,i}(\mathbf{z})$.

Data aggregator. Consider the example with a data aggregator where $\delta_i(\mathbf{z}) = b_i\phi(A)$ with $A = \int a_j z_j dj$. Suppose $a_i, b_i > 0$ for all i, $\phi(\cdot)$ is an increasing function, and the data policy that restricts cross-market data usage reduces b_i at least in a positive measure of markets. Using $\frac{\partial \delta_j(\mathbf{z})}{\partial z_i} = a_i b_j \phi'(A)$, we rewrite the equilibrium sales equation (19) as

$$G_i(z_i) = a_i \phi'(A) B + b_i \phi(A) \equiv \Delta_i(\mathbf{z}) \quad \text{where} \quad B = \int b_j z_j dj.$$

Given G_i is an increasing function under condition (21), we have $z_i = G_i^{-1}(a_i \phi'(A) B + b_i \phi(A))$, from which we can derive a system of equations on (A, B):

$$A = \int a_i G_i^{-1}(a_i \phi'(A) B + b_i \phi(A)) di; \quad B = \int b_i G_i^{-1}(a_i \phi'(A) B + b_i \phi(A)) di.$$

This converts the original infinite-dimensional problem into a two-dimensional problem. Once we solve (A, B), we can determine z_i and then all prices and qualities.

(i) Suppose first that $\phi(A)$ is weakly *convex*. Then $\Delta_i(\mathbf{z})$ is increasing in $\mathbf{z}$, and when data regulation reduces b_i in a positive measure of markets, it decreases $\Delta_i(\mathbf{z})$. Therefore, both conditions required in Proposition 2 hold, and so data regulation reduces z_i in each market and benefits every single-product firm.

(ii) Suppose instead that $\phi(A)$ is strictly *concave*. As in the above case, data regulation still decreases $\Delta_i(\mathbf{z})$. For the second condition that $\Delta_i(\mathbf{z})$ is increasing, we need that, for each $k \in \mathcal{I}$,

$$\frac{\partial \Delta_i(\mathbf{z})}{\partial z_k} = a_i a_k \phi''(A) B + (a_i b_k + b_i a_k) \phi'(A) > 0.$$

Given $B \leq \max_j \frac{b_j}{a_j} \times A$ and $\phi''(A) < 0$, a sufficient condition for the above inequality to hold is that, for any i, k,

$$\frac{\max_j \frac{b_j}{a_j}}{\frac{b_k}{a_k} + \frac{b_i}{a_i}} A \phi''(A) + \phi'(A) > 0 \iff R(A) < \frac{\frac{b_k}{a_k} + \frac{b_i}{a_i}}{\max_j \frac{b_j}{a_j}},$$

where $R(A) \equiv -\frac{A\phi''(A)}{\phi'(A)}$ denotes the coefficient of relative risk aversion when we interpret $\phi(\cdot)$ as a utility function. In particular, when all markets are symmetric (so that $\frac{b_i}{a_i} = \frac{b_k}{a_k}$ for any $i \neq k$) in which case the right-hand side reaches its maximum, the above condition simplifies to $R(A) < 2$, which is equivalent to $A\phi(A)$ being strictly convex.

Symmetric markets. Consider the example with symmetric markets. Slightly abusing the notation, let $\delta(z) = \delta_i(\mathbf{z})$ when $z_i = z$ for all $i \in \mathcal{I}$. Then in the symmetric equilibrium, $\delta'(z) = \int \frac{\partial \delta_i(\mathbf{z})}{\partial z_j} dj$. Using $\frac{\partial \delta_i(\mathbf{z})}{\partial z_j} = \frac{\partial \delta_j(\mathbf{z})}{\partial z_i}$ from the symmetry, we can simplify (19) to

$$G(z) = z\delta'(z) + \delta(z) = [z\delta(z)]'.$$

It has a unique (and stable) solution if the left-hand side has a slope in z greater than the right-hand side. Given $g(z) \geq \frac{3}{f(0)} - 2\overline{\eta}$, a sufficient condition for that is

$$\frac{3}{f(0)} - 2\overline{\eta} > [z\delta(z)]''$$

for any $z \in [0, 1]$. If the change of $\delta(\cdot)$ increases (decreases) $[z\delta(z)]'$ for any z, then equilibrium z increases (decreases). Note that in this symmetric case, we have clear-cut comparative statics even when the right-hand side of (19) is decreasing, which is beyond the scope of Lemma 4.

Notice that when data regulation reduces $\delta(z)$, $\delta'(z)$ might *increase*, in which case $[z\delta(z)]'$ may become greater, reducing the sales of single-product firms. To illustrate this possibility, consider an example with $\delta(z) = 1 - e^{-\gamma z}$. When γ decreases, $\delta(z)$ decreases everywhere but it also becomes less concave in the sense that the coefficient of absolute risk aversion $-\frac{\delta''(z)}{\delta'(z)} = \gamma$ decreases. More precisely,

$$\frac{d[z\delta(z)]'}{d\gamma} = ze^{-\gamma z}(2 - \gamma z).$$

Given $z \leq 1$, this is always positive when $\gamma \leq 2$. However, once $\gamma > 2$, this is negative when $z > \frac{2}{\gamma}$. Hence, in the latter case, if the ecosystem already holds a sufficiently large market share, a decrease in γ (i.e., a reduction in the ecosystem's data-driven quality) actually increases its market share and so *harms* single-product firms. This is because $\delta'(z) = \gamma e^{-\gamma z}$ decreases in γ when both γ and z are high, so a smaller γ increases the data subsidy $z\delta'(z)$ and leads to a lower price offered by the ecosystem.

B.2.5 The Continuum Case with the Within-Market Data Effect

In the continuum setting studied in Section 4, we have ignored the within-market data effect because the amount of data generated by a single product is negligible relative to the data from all other markets combined. However, if we interpret the continuum setting as the limit of the discrete setting with $n \to \infty$, as discussed in Section B.1.3, the within-market effect does not necessarily vanish in the limit. Data from, say, market i may be much more useful for product i than data from other markets, so that its impact on quality improvement can be comparable in magnitude to that of the entire cross-market effect (which must remain bounded under Assumption 1).

In this section, we reintroduce the within-market data effect into the continuum setting. To do so, we require the data effect $\delta_i(\mathbf{z})$ to satisfy the following property:

the impact of z_i on δ_i is comparable in magnitude to the effect generated by a positive measure of $z_{j\neq i}$. One convenient setup is to assume that

$$\delta_i(\mathbf{z}) = w_i(z_i) + c_i(\mathbf{z}) \quad \text{with} \quad w_i(z_i) \equiv w_{e,i}(z_i) - w_{s,i}(1 - z_i), \tag{56}$$

where $w_{e,i}$ and $w_{s,i}$ are standard differentiable functions, capturing the within-market data effect for the ecosystem and single-product firm i respectively, and c_i is a Fréchet differentiable functional, capturing the cross-market effect for the ecosystem. In this additive setup, we aim to show that our equilibrium analysis carries over, and the policy implications remain largely unchanged.[43]

Assumption 1 is now equivalent to, for any i and z,

$$|w_i'(z_i)| + \int \left|\frac{\partial c_i(\mathbf{z})}{\partial z_j}\right| dj < \frac{1-\epsilon}{f_i(0)}.$$

Given this assumption, the analysis of the consumer choice game remains unchanged.

Single-product firm's problem. Using $p_{s,i} = F_i^{-1}(z_i) - v_{e,i} + q_{s,i} - \delta_i(\mathbf{z})$, we write single-product firm i's profit as

$$[F_i^{-1}(z_i) - v_{e,i} + q_{s,i} - \delta_i(\mathbf{z})](1 - z_i) - \frac{q_{s,i}^2}{2\eta_{s,i}}.$$

Given $c_i(\mathbf{z})$ in $\delta_i(\mathbf{z})$ can be treated as a constant, it is straightforward to derive $q_{s,i} = \eta_{s,i}(1 - z_i)$ and

$$p_{s,i} = (1 - z_i)[h_i(z_i) - w_i'(z_i)].$$

This is the standard pricing formula in a model of single-product competition with a within-market data effect.

To investigate the second-order condition, using the optimal $q_{s,i}$ we can write single-product firm i's profit as a function of z_i:

$$[F_i^{-1}(z_i) - v_{e,i} - \delta_i(\mathbf{z})](1 - z_i) + \frac{\eta_{s,i}}{2}(1 - z_i)^2.$$

This is concave in z_i if

$$\eta_{s,i} + [F_i^{-1}(z_i)(1 - z_i)]'' \leq [\delta_i(\mathbf{z})(1 - z_i)]'' = [w_i(z_i)(1 - z_i)]''$$

[43] More generally, we could consider a function $\delta_i(z_i, \mathbf{z})$ that is differentiable (in the usual sense) in z_i and Fréchet differentiable in $\mathbf{z}$. However, several techniques (e.g., the potential function approach) used in the proof of equilibrium existence and uniqueness do not extend to this more general setup.

where the equality is because $c_i(\mathbf{z})$ in $\delta_i(\mathbf{z})$ is independent of z_i. This is true for any z_i if

$$\eta_{s,i} - \min_{z_i\in[0,1]}[w_i(z_i)(1-z_i)]'' \leq \min_{z_i\in[0,1]} h_i(z_i)[2-\sigma_i(z_i)].$$

Note that single-product firm i's profit is now

$$\pi_{s,i} = (1-z_i)^2\left[h_i(z_i) - w_i'(z_i) - \frac{\eta_{s,i}}{2}\right].$$

Using $\sigma_i(z) = (1-z)\frac{h_i'(z)}{h_i(z)}$, one can check that it decreases in z_i if and only if

$$\eta_{s,i} - [w_i(z_i)(1-z_i)]'' \leq h_i(z_i)[2-\sigma_i(z_i)].$$

This is guaranteed by the above second-order condition, just as in the case without the within-market effect.

Ecosystem's problem. The ecosystem's problem can be dealt with using the same approach as in Section B.2.2 without the within-market effect. We have $q_{e,i} = \eta_{e,i}z_i$ and

$$p_{e,i} = z_i[h_i(z_i) - w_i'(z_i)] - \int z_j \frac{\partial c_j(\mathbf{z})}{\partial z_i} dj.$$

The second-order condition remains unchanged provided that we adjust λ defined in footnote 25 accordingly by using the new $\delta_i(\mathbf{z})$. More precisely, now $\lambda = \max_{i,z_i}[z_i w_i(z_i)]'' + \lambda_c$, where λ_c is defined in footnote 25 with $\delta_i(\mathbf{z}) = c_i(\mathbf{z})$. Therefore, the ecosystem's second-order condition is, for any i,

$$\eta_{e,i} + \lambda_c + \max_{i,z}[zw_i(z)]'' \leq \min_{z\in[0,1]} h_i(z)[2-\sigma_i(z)].$$

Equilibrium existence and uniqueness. The system of equations for equilibrium sales now become

$$G_i(z_i) = z_i w_i'(z_i) + \int z_j \frac{\partial c_j(\mathbf{z})}{\partial z_i} dj - (1-z_i)w_i'(z_i) + \delta_i(\mathbf{z}).$$

Here the two terms involving $w_i'(z_i)$ are from the impact of the within-market effect on pricing. Using $\delta_i(\mathbf{z}) = w_i(z_i) + c_i(\mathbf{z})$, we can rewrite the above as

$$G_i(z_i) = w_i(z_i) + (2z_i - 1)w_i'(z_i) + \int z_j \frac{\partial c_j(\mathbf{z})}{\partial z_i} dj + c_i(\mathbf{z}) \equiv \Delta_i(\mathbf{z}). \tag{57}$$

The conditions for an interior solution take the same form as in Proposition 1 but with the newly defined $\Delta_i(\mathbf{z})$.

To find the condition for uniqueness of the solution, we adopt the same potential function approach as in the case without any within-market data effects, and define

$$\Gamma(\mathbf{z}) = \int \left\{ \int_0^{z_i} \lambda_c t_i - G_i(t_i) + w_i(t_i) + (2t_i - 1)w_i'(t_i)dt_i \right\} di + \left(\theta(\mathbf{z}) - \frac{\lambda_c}{2} \int z_i^2 di \right),$$

where $\theta(\mathbf{z}) = \int z_j c_j(\mathbf{z})dj$ and λ_c is the largest eigenvalue of $\theta(\mathbf{z})$ as defined in footnote 25. The second portion is concave by definition, and the first portion is concave as well if

$$\lambda_c - g_i(z_i) + [w_i(z_i) + (2z_i - 1)w_i'(z_i)]' \leq 0$$

for any $z_i \in (0, 1)$. Using $g_i(z_i) \geq \frac{3}{f_i(0)} - 2\bar{\eta}_i$, we derive a sufficient condition for the concavity of $\Gamma(\mathbf{z})$: for any i,

$$2\bar{\eta}_i + \lambda_c + \max_z [w_i(z) + (2z - 1)w_i'(z)]' < \frac{3}{f_i(0)}. \tag{58}$$

Policy implications. Using the newly defined $\Delta_i(\mathbf{z})$, our comparative statics with respect to the cross-market data effect carry over. More precisely, Proposition 2 continues to hold, as long as the condition for the consumer-surplus result is revised to

$$\eta_{s,i} > [1 + z_i - \sigma_i(z_i)]h_i(z_i) + [(1 - z_i)w_i'(z_i)]' - w_{s,i}'(1 - z_i)$$

where the w terms arise because $p_{s,i} = (1 - z_i)[h_i(z_i) - w_i'(z_i)]$ and the expression for consumer surplus now also includes the additional term $w_{s,i}(1 - z_i)$.

The impact of cross-market data sharing from the ecosystem to single-product firms also remains unchanged, i.e., Proposition 3 is still valid. This is simply because, given our additive setting, the within-market-effect terms in the equilibrium sales equation (57) are entirely separate from the cross-market-effect terms.[44] Similarly, the impact of introducing a data cooperative also remains the same, i.e., Proposition 4 still holds.

Within-market vs cross-market data effect. Our policy discussion has focused on the effects of cross-market data. We note that restricting within-market data usage is often less feasible than restricting cross-market data usage, and in the case of a data cooperative what single-product firms exchange is cross-market data. The within-market data effect is also more standard, analogous to network effects in competition. In our setup, the within-market and cross-market data effects can influence equilibrium

[44] In the analysis of the incentive to accept the compensation scheme, we now need $G_i(z_i) - [w_i(z_i) + (2z_i - 1)w_i'(z_i)]$ to be increasing in z_i, and this is ensured by (58) given $\lambda_c > 0$ when the cross-market effect is positive.

sales in different ways. This distinction is most transparent in the case of linear data effects. Slightly abusing notation, suppose the within-market data effect is $w_i z_i$ (where w_i is now a constant), while the cross-market data effect for product i is $\int c_{ij} z_j dj$. The equilibrium sales equation (57) then becomes

$$G_i(z_i) = (3z_i - 1)w_i + \int (c_{ji} + c_{ij}) z_j dj.$$

Reducing the c_{ij}'s necessarily lowers z_i, whereas reducing w_i lowers z_i if and only if initially $z_i > 1/3$.

B.3 Proofs and Details in the Tractable Discrete Case

Proof for Proposition 5. Suppose a symmetric interior equilibrium with $z_i = z \in (0, 1)$ for all i exists. The ecosystem's price for each product is

$$\begin{aligned} p_e &= zh(z) - z\left[\frac{\partial \delta_i(\mathbf{z})}{\partial z_i} + (n-1)\frac{\partial \delta_j(\mathbf{z})}{\partial z_i}\right] \\ &= z[L - (n-1)c], \end{aligned}$$

where we have used $h(z) = 2l$, $L = 2l - w$, and the linearity of the data effect in (31). Each single-product firm's price is

$$\begin{aligned} p_s &= (1-z)h(z) - (1-z)\left[\frac{\partial \delta_i(\mathbf{z})}{\partial z_i} + \frac{(n-1)\left[\frac{\partial \delta_j(\mathbf{z})}{\partial z_i}\right]^2}{h(z) - \frac{\partial \delta_i(\mathbf{z})}{\partial z_i} - (n-2)\frac{\partial \delta_j(\mathbf{z})}{\partial z_i}}\right] \\ &= (1-z)\left[L - (n-1)c\xi\right], \end{aligned}$$

where we have used μ_j^i derived in (52) and the notation ξ defined in (33).

Second-order conditions. In the symmetric-double-linear case, the ecosystem's second-order condition (9) simplifies to

$$\frac{\eta_e}{2} + (n-1)c \leq L \tag{59}$$

by using the facts that $\sigma_i(z) = 0$ and λ defined in Lemma 2 equals $2[w + (n-1)c]$.[45]

According to Lemma 3, with the double-linear specification a sufficient condition for single-product firms' second-order conditions to hold is $\eta_{s,i}\frac{\partial z_i}{\partial v_i} \leq 2$. When products

[45] One can readily check that the Hessian of $\sum z_i \delta_i(\mathbf{z})$ is a symmetric matrix with $2w$ being the diagonal entry and $2c$ the off-diagonal entry. Its largest eigenvalue is $2[w + (n-1)c]$ when $c \geq 0$.

are further symmetric, the J matrix in the proof of Lemma 1 simplifies to

$$J = \begin{bmatrix} L & \cdots & -c \\ \vdots & \ddots & \vdots \\ -c & \cdots & L \end{bmatrix}.$$

Then

$$\frac{\partial z_i}{\partial v_i} = \left(J^{-1}\right)_{ii} = \frac{|J_{-i}|}{|J|} = \frac{L-(n-2)c}{(L+c)(L-(n-1)c)},$$

where J_{-i} is the submatrix of J after removing the i_{th} row and column and the last equality used the formula $|A| = (x-y)^{n-1}[x+(n-1)y]$ when A is a symmetric $n \times n$ matrix with diagonal entry x and off-diagonal entry y. Using the ξ notation, we can write the condition $\eta_s \frac{\partial z_i}{\partial v_i} \leq 2$ as

$$\frac{\eta_s}{2(1-\xi)} - c \leq L. \tag{60}$$

Using the above expressions for p_e and p_s, we can write the equilibrium sales equation (13) as:

$$G(z) = w(2z-1) + (n-1)c[z - \xi(1-z)] + \delta(\mathbf{z}),$$

where

$$G(z) = F^{-1}(z) - [2\bar{\eta}z - \eta_s + (1-2z)h(z)] = 3l(2z-1) - 2\bar{\eta}z + \eta_s.$$

With the linear form of $\delta(\mathbf{z})$, it further simplifies to

$$G(z) = w(3z-1) - w_s + (n-1)c[2z - \xi(1-z)], \tag{61}$$

which is a linear equation and has a unique solution z defined in Proposition 5. For this solution to be stable, we need $g(z)$ to be greater than the slope of the right-hand side of (61) in z, i.e.,

$$3L - 2\bar{\eta} > (n-1)c(2+\xi). \tag{62}$$

Conditional on that, $z \in (0,1)$ is equivalent to $Z_s, Z_e > 0$ or the condition (35). Finally, note that (35) implies (62).

□

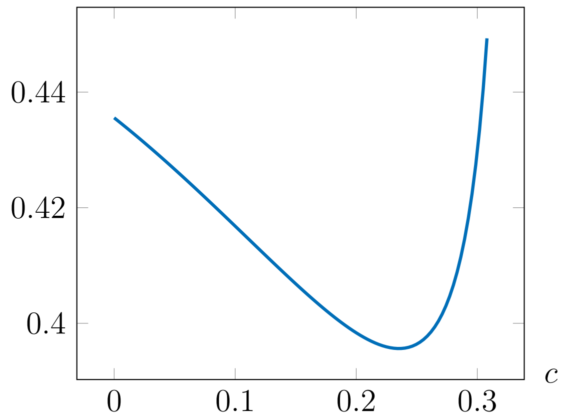


Figure 6: The impact of restricting cross-market data usage on single-product firms' profit in the discrete symmetric-double-linear case
($n = 2, l = 1, w_e = w_s = 0.05, \eta_e = 2.15$ and $\eta_s = 2.8$)

Data regulation and profit. Figure 6 reports an example in which a policy that reduces c can either increase or decrease single-product firms' profit, depending on the initial value of c (which affects the initial value of z).

As claimed in the main text, data regulation that reduces c cannot harm both the ecosystem and single-product firms.

Claim 5. *In the discrete symmetric-double-linear case, fixing all other parameters,* $\Pi_e(c_1) < \Pi_e(c_2)$ *and* $\pi_s(c_1) < \pi_s(c_2)$ *cannot hold simultaneously for* $0 \le c_1 < c_2$.

Proof. Recall that

$$\Pi_e(c) = z(c)^2 \left[L - (n-1)c - \frac{\eta_e}{2}\right],$$

and

$$\pi_s(c) = (1 - z(c))^2 \left[L - (n-1)c\xi - \frac{\eta_s}{2}\right].$$

where we have made the dependence of z on c explicit (fixing all other parameters). Note that the square-bracketed terms in each profit expression are decreasing in c given ξ increases in c. Hence, if we have $\Pi_e(c_1) < \Pi_e(c_2)$, we must have $z(c_1) < z(c_2)$. But then necessarily $\pi_s(c_1) > \pi_s(c_2)$. □

Proof for the consumer-surplus result. Here we prove the claim in the main text when $c \to 0$ or $n \to \infty$. Note that

$$\frac{dV}{dc} = \frac{\partial V}{\partial z}\frac{dz}{dc} + (1-z)(n-1)\frac{d(c\xi)}{dc}.$$

Using

$$\frac{\partial V}{\partial z} = 2lz - [\eta_s + w_s + (n-1)c\xi - L], \quad \frac{dz}{dc} = (n-1)\frac{2z - (1-z)[2\xi + (n-2)\xi^2]}{3L - 2\bar{\eta} - (n-1)c(2+\xi)},$$

and

$$\frac{d(c\xi)}{dc} = 2\xi + (n-2)\xi^2,$$

one can check that

$$\frac{dV}{dc} < 0 \iff 2z\frac{\frac{\partial V}{\partial z}}{D} + (1-z)[2\xi + (n-2)\xi^2]\left(1 - \frac{\frac{\partial V}{\partial z}}{D}\right) < 0$$

where $D = 3L - 2\bar{\eta} - (n-1)c(2+\xi) > 0$ under condition (35). When $c \to 0$, we have $\xi \to 0$ and so the above condition holds if and only if $\frac{\partial V}{\partial z} < 0$. When $n \to \infty$ (in which case $c \to 0$ as well), we also have $\xi, (n-2)\xi^2 \to 0$ under Assumption 1, so again the above condition holds if $\frac{\partial V}{\partial z} < 0$. Moreover, notice that in both limit cases $(n-1)c\xi \to 0$ under Assumption 1, and so $\frac{\partial V}{\partial z} < 0$ if and only if $z < \frac{1}{2l}(\eta_s + w_s - L)$.[46] Since z decreases in η_s, this condition holds if η_s is above a threshold.

Data sharing. In the continuum case, we saw that free data sharing has the same impact on firms as restricting cross-market data usage, and is more beneficial for consumers. This remains true if only cross-market data is shared. Specifically, with cross-market data sharing, the ecosystem's relative data advantage in market i becomes

$$\delta_i(\mathbf{z}) = w_e z_i + (c_e - c_s)\sum_{j \neq i} z_j - w_s(1 - z_i)$$

where c_e and c_s are the cross-market data effect coefficients respectively for the ecosystem and the single-product firms. It is then obvious that increasing c_s has the same impact on the equilibrium as reducing c_e.

On the other hand, our analysis of compensated data sharing in the continuum case does not extend for two reasons: (i) Since a single-product firm's choice affects sales in all the markets, any meaningful compensation scheme will affect single-product firms' pricing decisions, which will complicate the analysis of whether each single-product firm has an incentive to accept the scheme. (ii) Due to the feedback effect in p_s, it is hard to find a simple compensation scheme that will induce the same sales as absent data sharing.

[46]This condition is nonempty. When $c \to 0$, $z \to 0$ if $L + l \to w_s + \eta_s$, under which the right-hand side is 1/2.

Data cooperative. Given a data cooperative (with free data exchange) among all single-product firms, we have

$$\delta_i(\mathbf{z}) = w_e z_i + c_e \sum_{j \neq i} z_j - w_s(1 - z_i) - c_s \sum_{j \neq i}(1 - z_j),$$

where c_s is now the cross-market data effect within the cooperative. Strengthening the data cooperative (i.e., increasing c_s) now not only induces the ecosystem to price more aggressively, as in the continuum case, but also induces single-product firms to lower their prices due to the feedback effect. As a result, the data cooperative is more likely to increase single-product firms' sales. More precisely, using $c \equiv c_e + c_s$, the equilibrium sales equation is

$$G(z) = w(3z - 1) - w_s + (n - 1)c[2z - \xi(1 - z)] - (n - 1)c_s.$$

The marginal impact of an increase of c_s on the right-hand side is $n - 1$ times

$$2z - (1 - z)[2\xi + (n - 2)\xi^2] - 1.$$

This is negative—in which case introducing the data cooperative increases each single-product firm's sales—if and only if

$$1 - z > \frac{1}{2 + 2\xi + (n - 2)\xi^2}.$$

In the limit case with $n \to \infty$, the right-hand side goes to $1/2$, consistent with Corollary 1 in the continuum case. With a finite n, however, the right-hand side must be less than $1/2$. This confirms that, compared to the continuum case, it is easier in the discrete case for a data cooperative to raise single-product firms' sales. However, this does not necessarily imply that a data cooperative is more likely to benefit single-product firms in the discrete case, since an increase in c_s also induces them to offer lower prices, potentially offsetting the benefits of selling more.