CESifo WORKING PAPERS

10963 2024

February 2024

The Strategic Value of Data Sharing in Interdependent Markets

Hemant Bhargava, Antoine Dubus, David Ronayne, Shiva Shekhar

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10963*

The Strategic Value of Data Sharing in Interdependent Markets

Abstract

Large, generalist, technology firms—so-called "big-tech" firms—powerful in their primary market, routinely enter secondary markets consisting of specialist firms. Naturally, one might expect a specialist firm to be fiercely protective of its data as a way to maintain its market position in the secondary market. Counter to this intuition, we demonstrate that a specialist firm willingly shares its market data with an intruding tech generalist. We do so by developing a model of cross-market competition in which data collected via consumer usage in each market is a factor of product quality in both markets. We show that a specialist firm shares its data to strategically create *co-dependence* between the two firms, thereby softening competition and transforming the generalist firm from a traditional *competitor* into a *co-opetitor*. For the generalist intruder, data from the specialist firm substitute for its own investments in product quality in the secondary market. As such, the act of sharing data makes the intruder a stakeholder in the valuable data collected by the specialist, and consequently in the specialist's continued success. Moreover, while the firms benefit from data sharing, consumers can be worse off from the weaker price competition and lower investments in innovation. Our results have managerial and policy implications, notably on account of backlash against data collection and the market power of big tech firms.

Keywords: data-driven quality improvements, externalities, co-opetition, data sharing.

Hemant Bhargava
University of California, Davis / CA / USA
hemantb@ucdavis.edu

Antoine Dubus
ETH Zurich / Switzerland
antoine1dubus@gmail.com

David Ronayne
European School of Management and Technology (ESMT Berlin) / Germany
david.ronayne@esmt.org

Shiva Shekhar
Tilburg School of Economics and Management
Tilburg University / The Netherlands
s.shekhar_1@tilburguniversity.edu

February 17, 2024
We would like to thank participants at the 2022 Workshop on Information Systems and Economics, the Paris Digital Economics Seminar, the 2022 Workshop on Platforms at the National University of Singapore, the Joint Research Center Economic Seminar, and the 2023 SSES Annual Conference. Ronayne is grateful for support from Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119).

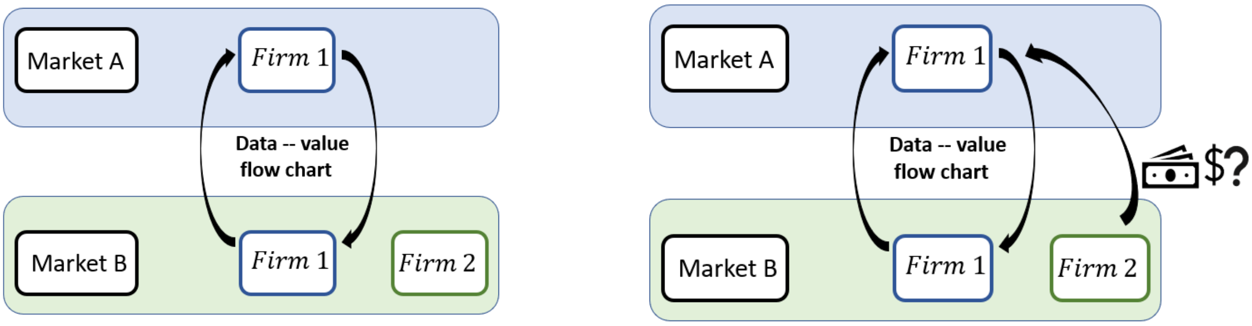
1. Introduction

So-called "big tech" firms such as Google, Facebook, Apple, Amazon and Microsoft have become the corporate giants of recent years. These firms are data powerhouses, collecting and leveraging massive amounts of user data through the use of sophisticated algorithms, and transforming data-driven insights into products and services which generate higher per-user monetization and revenues. Their ever-present hunger for data, among other reasons, often leads these firms into adjacent markets. For instance, Google has combined its dominance in search with moves into areas such as Android hardware (such as smartphones and tablets), Chromebook laptops, wearables (including smartwatches and fitness devices), home automation (e.g., hubs, routers, sensors), and autonomous driving. Similarly, Amazon has gone beyond even its "Superstore" vision by extending its prowess into cloud computing (AWS), entertainment (Amazon Video), game streaming (Twitch), and logistics. Such cross-market moves have extended the reach of big tech firms well beyond their original markets.

One advantage big tech firms have is their ability to leverage data across markets and thereby create better products. For instance, a search engine can fine-tune its search results and sponsored advertising algorithms by using sales and conversion data from a retail operation; conversely, it can leverage users' search data to hone its retail tactics.[1] Similarly, data about driver actions captured by an autonomous vehicle can trigger improvements in mapping data, and mapping data are of course useful for driving.[2] Thus, cross-market data sharing creates a virtuous cycle (see Fig. 1a), as more sales or usage in one market generates more data, which provides the basis for improved analytics and algorithmic learning, leading to better offerings in both markets. This "data-driven network effect" (Argenton & Prüfer 2012, Gregory et al. 2021, Prüfer & Schottmüller 2021) has taken a central role in the debate on the regulation of big tech (Cennamo & Sokol 2021, European Commission 2020*a*,*b*, Krämer & Schnurr 2022, Parker et al. 2021). With their prowess in artificial intelligence and machine learning techniques fuelled by abundant data, such firms derive a competitive advantage from their ability to leverage data

[1] See Klein et al. (2022) for a discussion of how data is a key input in improving the quality of search results.
[2] E.g., "Waze Could Be Google's Ace in the Hole in a Self-Driving Car War With Uber," last accessed Aug 21, 2023: vox.com/2015/12/18/11621572/waze-could-be-googles-ace-in-the-hole-in-a-self-driving-car-war-with

between their original and secondary markets. In recent work, Lei et al. (2023) provide evidence from a large-scale experiment that showcases the complementary value of cross-market data.[3]



(a) Data-value flows across markets.

(b) Is sharing data profitable?

Figure 1: Data-Sharing between Firms across Markets

In light of the data-driven network effect advantage that big tech firms enjoy when entering secondary markets, this paper examines what tactic incumbent "specialist" firms (which only operate in the secondary market) should adopt with regard to proprietary data in their markets. To facilitate exposition, the generalist big tech firm, which operates in a primary market A but can also enter a secondary market B, is labeled firm 1. Market B features an incumbent firm 2 that only operates in this market (see Fig. 1a). In light of the generalist firm's ability to harness data-driven network effects to improve its competitive position in market B, one would expect incumbent firm 2 to be highly protective of its own proprietary data market B. If firm 1 could get firm 2's market B data, it would improve both its market A and market B products. Yet, we observe contrary strategy in practice. For example, the smartwatch company Mobvoi TicWatch, whose Pro 5 smartwatch competes with Google's Pixel and Sense smartwatches, requires that users consent to Google Cloud Sync terms of service, including granting Google access to the data collected by Mobvoi.[4] More generally, reports by the European Commission

[3]Specifically, they show that a query autocomplete company's click-through rate rises about 5% with access to (another firm's) search engine's data. The authors also provide an up-to-date review of several other empirical studies documenting the benefits of cross-market data sharing.

[4]See, for example, "Mobvoi TicWatch Pro 5 review: timing is everything," last accessed, Nov 14, 2023 https://www.theverge.com/23733359/mobvoi-ticwatch-pro-5-review-wear-os-3-smartwatch-wearables. Other examples include home devices such as thermostats, cameras and sensors that capture data about the lifestyles of their users. Those data complement the health-related data collected by wearables such as smart watches, resulting in finer sport, fitness, and activities recommendations. Conversely, the data collected by wearables can be used to optimize home devices' design and performance.

(2018) point to a large and sharply increasing practice of B2B data-sharing.[5,6] They also found 90% of companies sharing data did so with others within the same broad business sector.

These dynamics inspire us to ask the following questions, depicted in Fig. 1b: Does a specialist firm in fact have an incentive to share data with a generalist rival? If so, when and why might such a counter-intuitive strategy make sense? What is the impact on profits and consumers, and, ultimately, entry decisions? Our main contribution is to identify a novel strategic rationale for why specialist firms may want to share data with their generalist rivals, *for free*. Our analysis and the intuition for this result is as follows. First, we confirm that data sharing by firm 2 (of its market B data to firm 1) indeed is a competitive gift to firm 1; doing so enables firm 1 to improve its product quality in market B. Second, however, we show that receipt of this gift makes firm 1 *reliant* on firm 2, causing it to view firm 2 as a *co-opetitor* in market 1 rather than a traditional *competitor*. The logic is that when the specialist firm 2 enjoys greater demand, then it can provide more data to firm 1, increasing firm 1 profit. Moreover, this profit comes at a lower cost than firm 1 would incur if it made quality improvements only through direct investments. Consequently, it *wants* firm 2's market data, and therefore has a stake in firm 2's continued success in market B. This causes firm 1 to scale back its own entry into B, mitigating the competitive pressure on firm 2.

In support of this intuition, the European Commission (2018) points to B2B agreements having "different conditions to share data [...], including for free." There is also an emerging industry of firms that facilitate and support such B2B data sharing.[7] Recognizing the importance of this trend, European policy makers have even launched several public initiatives to foster data sharing between firms and improve the efficiency of data-intensive companies, in particular for firms in the same industry.[8] Our results suggest that encouraging data sharing among rival firms is in fact a way for policy makers to help small firms in the face of tech giants.

[5]Of the companies surveyed, 37% share data, and 14% share more than 50% of the data they generate, including 2% of companies that generate more than 1PB/month—figures the report emphasizes are rapidly increasing.

[6]Speaking to the expected growth in sharing, the European Commission (2020*c*) will, over 2021-27, invest in European data spaces and federated cloud infrastructures, as part of the European Strategy for Data.

[7]For example, StartUs Insights: https://www.startus-insights.com/innovators-guide/discover-5-top-data-sharing-startups-scaleups/, last accessed Jan 19, 2023.

[8]E.g., project GAIA X: https://www.bmwk.de/Redaktion/EN/Dossier/gaia-x.html, last accessed Feb 13, 2024.

Despite the apparent pro-competitive aspects of data sharing, there is a need for a nuanced view of this practice. Data sharing from a specialist to a generalist big tech firm is a win-win for both firms (unless a very aggressive generalist poses an existential threat to the specialist), however it might be detrimental for customers. When the firms engage in co-opetition, the less intense competition can make customers worse off through less innovation and weaker price competition. Widening the scope of consideration, long-term competitive dynamics can be seen to weaken or strengthen these observations. On the one hand, if data fuels long-term innovation, data sharing may help the generalist firm to exclude its specialist rival from the market in the long run. On the other hand, as the specialist firm invests more in innovation when data are shared, this may also help its long-term growth. We discuss the possible interactions between the mechanism we identify and other effects in a broader context in Section 6.

Regarding policy and welfare, several initiatives have been recently implemented by regulators to foster data sharing practices between firms (see for instance the GAIA-X initiative in Europe).[9] Our results send a cautionary note to policy makers when it comes to the pro-competitive effects of data and the benefits of data sharing practices for consumers. As noted above, although data sharing enhances value creation, it can make consumers worse off when the reduction in competition overpowers the value creation effect of data sharing. Hence, an unintended anti-competitive effect may arise when firms implement such data-sharing practices. This suggests that there is no such thing as a one-size-fits-all mandatory data sharing policy. Rather, our results suggest that policy makers should be especially cautious when it comes to data sharing in interrelated markets, in which case a small firm sharing its data with a competing conglomerate operating in interrelated markets may be of concern. Instead, this negative effect of data sharing on consumers does not take place when markets are unrelated, suggesting that data sharing practices may create value and increase welfare.

[9]See Footnote 8 for a link to more information on that initiative.

2. Literature

Information sharing. Scholars and practitioners have examined information sharing between firms for over 50 years. A large theoretical literature analyzes whether competing firms have incentives to exchange information about market characteristics such as consumer demand and production costs (Raith 1996). Influential papers include Vives (1984) and Gal-Or (1985), which characterize when it is profitable for duopolists to share information about uncertain demand. With recent advancements in technologies such as machine learning and AI, consumer-level data are important inputs in digital markets, the sharing (or protection) of which has become a key subject for analysis. For example, Jones & Tonetti (2020) show that dominant firms may choose to hoard their datasets to preserve their competitive advantage. In contrast, Choe et al. (2022) study the incentives of firms to share data in a context where data are used to price discriminate. They show that a firm endowed with data on consumers would be willing to give certain data to a data-less competitor, to soften price competition. In a similar spirit, Huang et al. (2020) show that sharing intellectual property can soften competition between firms when learning costs are sufficiently high. In contrast to these studies, we develop a model in which data are valuable across markets and firms are ex-ante asymmetric: we study the incentive of specialist (single-market) firms to share data with their generalist (multi-market) rivals. We find that specialist firms may have incentives to share data with their generalist rivals as a strategic device to lower competitive intensity.

Partial ownership. By sharing its data, a specialist firm gives its competitor access to its assets, and for this reason our paper relates to the literature on partial ownership (see, e.g., Ederer & Pellegrino 2022, Gilo et al. 2006, Hunold & Shekhar 2022, O'Brien & Salop 1999, Antón et al. 2023). The competition reduction induced by partial ownership has received the attention of policy makers, and appropriate regulations have been formulated.

Our contribution is to show that, in contrast to partial ownership effects, both firms have higher profits when data are shared via arrangements free of other obligations. As such, there is no loss for the specialist firm in sharing data *per se*, while partial ownership usually requires a firm to give away part of its control over its activities. Moreover, data sharing may not bear the contractual complexity of partial ownership and so may be particularly ripe for adoption by data management teams. In particular, partial ownership deals are costly to establish and reverse, while data sharing can be more easily turned on or shut down.

From a regulatory perspective, while policymakers are usually wary of partial ownership practices, B2B data sharing has thus far seemingly slipped through the anti-trust dragnet. Instead, it has had mostly the positive (surplus-generating) aspects highlighted, in places leading to an encouragement of the practice in general.[10]

Data-driven network effects. We contribute to the emerging literature analyzing the economic impacts of data-driven network effects (Gregory et al. 2021). Argenton & Prüfer (2012) consider those effects in the context of search engines and find they can lead to market tipping. Schaefer et al. (2018) also consider the search engine market and use real search engine query logs to empirically investigate the quality improvements from such effects. Prüfer & Schottmüller (2021) study the investment incentives of competing firms under data-driven network effects in a dynamic setting, including when data in one market can be leveraged in another. Unlike Prüfer & Schottmüller (2021), we model data-driven network effects as enhancing user experience. This allows us to study the interplay between competitive strategies and innovation decisions. The cross-market interaction encourages more innovation in the primary and secondary market by the generalist firm than if the two markets were not connected or regulated. Further, we allow competing firms to choose how much to invest directly in innovation via costly methods such as R&D.

Cross-market externalities. Our work also contributes to the literature on cross-market externalities where the positive impact of corporate diversification on firms' profitability has been empirically established (Berger & Ofek 1995, Graham et al. 2002, Lang & Stulz 1994, Lins & Servaes 1999). The rationale for such a strategy is the resulting synergies which may

[10]These policy recommendations include the European Data Governance Act and the Data Act (which can be accessed respectively at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113; and https://digital-strategy.ec.europa.eu/en/policies/data-governance-act).

stem from economies of scope in production and the uses of common distribution channels for different products within a firm, such as the Apple Store for Apple's physical goods, or Google Play for mobile applications (Hill & Hoskisson 1987), or from innovation spillovers across related products (Baysinger & Hoskisson 1989).[11] In a similar spirit, we model competition between a large multi-market firm, and a smaller single-market competitor.[12] We model synergies as the value generated from relevant data collected in other markets, which activates a virtuous cycle of data-driven network effects. There, we contribute by studying the incentives and effects of a specialist firm to share its data with a generalist competitor via inter-market synergies.

3. Model

We study a game-theoretic model with two firms, $i = 1, 2$, and markets for two goods, A and B, with production costs set to zero for simplicity. Firm 1 is a monopolist in market A. Firm 2 operates in market B, which firm 1 enters. We model firm 2's decision to share data with firm 1. Each firm also chooses how much to produce and how much to invest in product quality.[13]

Market A is governed by an (inverse) demand function $P_A = \mathcal{A} - \beta_A q_A$ (where $\mathcal{A}$ is base quality, P is price, and q is quantity). This extends the classical (linear) form in the following ways, representing two mechanisms for firm 1 to improve the value delivered to its users in market A.[14] First, firm 1 can shift the demand curve up by v_A via direct investments in innovation or operational expenditures such as customer support or service infrastructure, at cost $I(v_A) = \frac{v_A^2}{2}$. Second, it can shift the demand curve up by using data from market B. This

[11] Additionally, Gomes & Livdan (2004) argue that diversification allows corporate firms to undertake potentially high-reward risky projects while securing a steady cash flow from other, more stable markets.

[12] This market configuration is supported by empirical evidence of product diversification by digital giants, a strategy that has been extensively analyzed by the economics and management literature, back to Hill & Hoskisson (1987) and Shaked & Sutton (1990).

[13] Although the model is setup with 2 firms, our results are reinforced when more than 2 firms compete in market B. This is because competition in market B makes it more likely that small firms will share data with their multi-market rival. In favor of being conservative by stacking the odds against incentives to share data, we shut down this competitive effect to solely focus on our main mechanism. For our narrative, we can think of firm 1 as a big-tech company such as Google, which decides to enter market B where firm 2 is a smaller incumbent. By operating in both markets, firm 1 benefits from cross-market data externalities.

[14] We show our results also follow under price competition (see Section 5.1) and in the presence of various other extensions, which we house in Section 5 and our online-only Appendix C.

includes data generated by its own sales in market B, q_1, and, if firm 2 shares, data from firm 2's sales in B, q_2. Let the indicator variable Φ be 1 when B shares its data with firm 1 and 0 otherwise. Then total data available to firm 1 in market A is $q_1 + \Phi q_2$. Additionally, let $\theta > 0$ be the rate at which a unit of data translates into higher willingness to pay of consumers.[15] Incorporating those features, firm 1's demand in market A is:

$$P_A(v_A, q_1, \Phi q_2, q_A) = \mathcal{A} + \underbrace{v_A}_{\substack{\text{quality increase}\\ \text{by investing}}} + \underbrace{\theta(q_1 + \Phi q_2)}_{\substack{\text{data}\\ \text{advantage}}} - \beta_A q_A. \tag{1}$$

The last term is the standard (linear) inverse relationship between price and sales. The intercept term is $\mathcal{A} = \alpha + \beta_A/2$. One can interpret $\alpha \geq 0$ and $\beta_A > 0$ as the average and spread of willingness to pay, respectively.[16]

Market B operates under an inverse demand function $\mathcal{B} - \beta_B Q_B$ (where Q_B is the sum of outputs q_1 and q_2 of firms 1 and 2), which is similarly adjusted for quality-related investments and, only for firm 1, an additional quality increment due to its data advantage.[17] Firm 1 can use the data it generates from its sales in market A, q_A, to improve its product in B, shifting demand by θq_A.[18] The resulting (inverse) demand functions for each firm in market B are:

$$P_1(v_1, q_A, Q_B) = \mathcal{B} + v_1 + \underbrace{\theta q_A}_{\substack{\text{data}\\ \text{advantage}}} - \beta_B Q_B, \quad P_2(v_2, Q_B) = \mathcal{B} + v_2 - \beta_B Q_B, \tag{2}$$

where v_i for $i = 1, 2$ is the quality increase by investing. Profits of firms 1 and 2 are, respectively:

$$\Pi_1 = \underbrace{P_A(\cdot)q_A - I(v_A)}_{\text{Profits from market A}} + \underbrace{P_1(\cdot)q_1 - I(v_1)}_{\text{Profits from market B}}, \quad \Pi_2 = P_2(\cdot)q_2 - I(v_2). \tag{3}$$

The timing of the game is as follows. At stage 1, firm 2 decides whether to share data with

[15] For example, data from a search engine about users' queries and clicks on fitness-related searches could lead to better dashboards and improves the default features in a fitness app which increases user willingness to pay.

[16] We provide a corresponding microfoundation for this interpretation in online Appendix C.

[17] In an extension, we show that our results hold in the presence of same-side network externalities (see Section 5.2). We also conduct this exercise in Appendix C, but in the presence of price competition.

[18] In the baseline framework, we consider identical θ identical from market A to B and reciprocally and no intra-market externalities. We provide extensions without those assumptions in Section 5.

firm 1. At stage 2, both firms choose their level of expenditures to determine v_A, v_1 and v_2 in the markets they operate in. At stage 3, firms set outputs and simultaneously compete to serve consumers. Last, demand and profits are realized. We seek Subgame Perfect Nash Equilibria and solve the game backwards. We present a summary of notations below in Table 1.

Table 1: Notation Guide

Variable	Interpretation
i	Index to represent firms $i = 1, 2$.
P_A	Inverse demand function of firm 1 in market A.
P_i	Inverse demand function of firm $i \in \{1, 2\}$ in market B.
$\mathcal{A}, \mathcal{B}$	Base quality of products in markets A and B.
θ	Productivity of data-driven network effect.
$I(v)$	Investment cost to reach innovation level v.
Π_i	Profit of firm $i \in \{1, 2\}$.
CS_m	Consumer surplus in market $m \in \{A, B\}$.
β_k	Output sensitivity of demand in markets k for $k \in \{A, B\}$.
q_A	Consumer demand in market A.
q_i	Consumer demand of firm i in market B.
v_A	Innovation effort by firm 1 in market A.
v_i	Innovation effort by firm i in market B for $i \in \{1, 2\}$.

The following regularity conditions ensure that there is an interior equilibrium solution (which is the non-trivial case because under corner solutions, demands become inelastic):[19]

- The data externality is not too strong: $\theta < \overline{\theta} \approx 0.353$.
- The demand intercept terms are not too high: $\mathcal{A}, \mathcal{B} < \frac{2(42-88\theta^2+43\theta^4-6\theta^6)}{(3-\theta^2)(12+\theta(4-\theta(19+2\theta-4\theta^2)))}$.
- The spread of the willingness to pay is sufficiently high: $\beta_A, \beta_B > \tilde{\beta} \approx 1.7$, and for increased tractability we set $\beta_A = \beta_B = 2$.

Data shared by firm 2 allows firm 1 to attract more demand and collect more data in market A. Hence, data sharing also increases the competitiveness of firm 1 in market B through enhanced cross-market data externalities from A to B, so that sharing data could have a negative impact on firm 2. Viewed in this way, our setup stacks the odds against firm 2 sharing data with firm 1, especially for free, as we confirm in Lemma 1.

[19] We discuss our assumptions and probe our model's robustness in Section 5 and Appendix C.

4. Analysis and Results

4.1. A benchmark case without entry

Before solving the model, consider a hypothetical benchmark case where firm 1 is present only in market A and does not enter market B at all, rendering the two markets separate and independent. In that setting, data sharing has no impact on firm 2, but it carries some advantage for firm 1. Hence, under these separate markets, firm 1 would be willing to pay some price (up to its gain from data sharing) while firm 2 would be willing to share data for any positive payment from firm 1. With this in mind, we return to the main setting where firm 1 enters market B, and we show that firm 2 would still be willing to share its data with firm 1, even for free, despite the direct competitive threat it faces in market B.

4.2. Main model

Output-setting stage. The first-order conditions for the two firms' output decisions are

$$\frac{\partial \Pi_1}{\partial q_k} = \underbrace{P_k(\cdot)}_{\text{Volume effect}} + \underbrace{\frac{\partial P_k(\cdot)}{\partial q_k} q_k}_{\text{Margin effect}} + \underbrace{\frac{\partial P_j(\cdot)}{\partial q_k} q_j}_{\substack{\text{Value increase in market B} \\ \text{from data collected in market A (+)}}} = 0, \text{ for } k \neq j \in \{A, 1\}, \quad (4)$$

$$\frac{\partial \Pi_2}{\partial q_2} = \underbrace{P_2(\cdot)}_{\text{Volume effect}} + \underbrace{\frac{\partial P_2(\cdot)}{\partial q_2} q_2}_{\text{Margin effect}} = 0. \quad (5)$$

Firms face the standard volume and margin trade-off. In addition, firm 1 benefits, in each market, from increased margins due to the value it creates with data from the other market.[20] Solving simultaneously, the above system yields the equilibrium outputs as functions of the sharing decision and quality improvement levels, denoted by $\hat{q}_A(v_A, v_1, v_2, \Phi)$, $\hat{q}_1(v_1, v_A, v_2, \Phi)$, $\hat{q}_2(v_2, v_1, v_A, \Phi)$ and $\hat{Q}_B(v_2, v_1, v_A, \Phi) = \hat{q}_1(\cdot) + \hat{q}_2(\cdot)$.[21] Intuitively, as quality improvements increase consumers' willingness to pay, firms produce more in the corresponding market, and

[20] An increase in output by firm 1 in market B enhances the margin on its sales in market A, thus making it profitable to expand output in market B. A similar intuition holds for firm 1's output strategy in market A.

[21] The terms we reference in the text are fully written out in Appendix A's proofs of Lemma 1 and Proposition 1.

so $\frac{\partial \hat{q}_1(\cdot)}{\partial v_1}, \frac{\partial \hat{q}_A(\cdot)}{\partial v_A}, \frac{\partial \hat{q}_2(\cdot)}{\partial v_2} > 0$. Because of the cross-market data externality, if firm 1 invests more in quality (increasing output, which generates more data) in one market, its product in the other market also improves, boosting demand, so firm 1 produces more there too, i.e., $\frac{\partial \hat{q}_A(\cdot)}{\partial v_1}, \frac{\partial \hat{q}_1(\cdot)}{\partial v_A} > 0$. In contrast, firms produce less when their rival invests more in quality. For example, if firm 1 invests more in market B, then it produces more in B, which increases the competitive pressure on firm 2 that responds by producing less because output choices are strategic substitutes. That is, we have the following relations: $\frac{\partial \hat{q}_A(\cdot)}{\partial v_2} \leq 0, \frac{\partial \hat{q}_1(\cdot)}{\partial v_2}, \frac{\partial \hat{q}_2(\cdot)}{\partial v_A}, \frac{\partial \hat{q}_2(\cdot)}{\partial v_1} < 0$.

In this output-setting phase, and holding investments constant, firm 2 should not share its data with firm 1.

Lemma 1 (**Data Sharing Without Investment Responses**). *Keeping investments constant, firm 2 has no incentive to share data with firm* 1*, i.e.,* $\Pi_2(0) - \Pi_2(1) > 0$.

Without responses in investment, data sharing by firm 2 returns to haunt it because it enhances firm 1's value proposition in market A and, via cross-market data network effects, strengthens its position in market B too. This in turn lowers both firm 2's output and profit, implying that it should not share its data (for free). This result lays some groundwork with which to view our contribution: Lemma 1 is reversed when one accounts for direct investments as another lever to improve product quality.

Innovation-setting stage. Substituting in the optimal output choices, we write the demand functions in terms of choices made at stages 1 and 2: $\hat{P}_A(v_A, v_1, v_2, \Phi)$, $\hat{P}_1(v_A, v_1, v_2, \Phi)$, and $\hat{P}_2(v_2, v_1, v_A, \Phi)$. Firms set innovation levels to maximize profits:

$$\max_{v_A, v_1} \hat{\Pi}_1(v_A, v_1, v_2, \Phi) = \sum_{k=1,A} \hat{P}_k(\cdot)\hat{q}_k(\cdot) - I(v_k), \ \max_{v_2} \hat{\Pi}_2(v_2, v_1, v_A, \Phi) = \hat{P}_2(\cdot)\hat{q}_2(\cdot) - I(v_2). \quad (6)$$

Applying the envelope theorem to the first-order conditions, we obtain the following system:

$$\frac{\partial \hat{\Pi}_1(\cdot)}{\partial v_k} = \underbrace{\underbrace{\frac{\partial P_k(\cdot)}{\partial v_k}\hat{q}_k(\cdot)}_{\text{Margin effect}} - \underbrace{\frac{\partial I(v_k)}{\partial v_k}}_{\text{Cost}}}_{\text{Direct effects}} + \underbrace{\underbrace{\frac{\partial P_1(\cdot)}{\partial Q_B}\frac{\partial \hat{q}_2(\cdot)}{\partial v_k}\hat{q}_1(\cdot)}_{\text{Competitive effect }(+)} + \underbrace{\overbrace{\frac{\partial P_A(\cdot)}{\partial q_2}}^{=\Phi}\frac{\partial \hat{q}_2(\cdot)}{\partial v_k}\hat{q}_A(\cdot)}_{\text{Data-sharing effect }(-)}}_{\text{Strategic effects }(?)} = 0,\ k \in \{A, 1\} \tag{7}$$

$$\frac{\partial \hat{\Pi}_2(\cdot)}{\partial v_2} = \underbrace{\underbrace{\frac{\partial P_2(\cdot)}{\partial v_2}\hat{q}_2(\cdot)}_{\text{Margin effect}} - \underbrace{\frac{\partial I(v_2)}{\partial v_2}}_{\text{Cost}}}_{\text{Direct effects}} + \underbrace{\frac{\partial P_2(\cdot)}{\partial Q_B}\frac{\partial \hat{q}_1(\cdot)}{\partial v_2}\hat{q}_2(\cdot)}_{\text{Competitive effect }(+)} = 0. \tag{8}$$

The terms labeled *direct effects* are two classic and opposing forces. A unit increase in quality increases consumers' willingness to pay and thus also the firm's *margins*, boosting profit, but also requires *costly* investment. The *strategic effects* account for the rival's reactions to quality improvements. Here, there are two opposing effects. First, the *competitive effect* represents the increased profitability for a firm in market B following its rival's scaling back in market B. The second is the *data-sharing effect*, which is triggered when firm 2 shares data ($\Phi = 1$). This effect represents the decreased profitability of firm 1 in market A caused by the reduction in the amount of data available to it in A following 2's contraction in market B (and the corresponding fall in data shared by firm 2). This effect dampens the strategic incentives of firm 1 to invest in quality improvement when firm 2 shares data ($\Phi = 1$). It is a manifestation of the transformation of the relationship of the firms from *competition* to *co-opetition.* The strength of this effect determines whether firm 1 innovates more in market B or in market A. Solving these first-order conditions simultaneously yields the equilibrium quality improvement levels as function of firm 2's sharing decision, $\Phi \in \{0, 1\}$, denoted by $v_A^\star(\Phi)$, $v_1^\star(\Phi)$ and $v_2^\star(\Phi)$. Substituting the optimal investment choices at stage 2 into the optimal outputs as functions of quality improvement at stage 3 yields the equilibrium outputs, $q_A^\star(\Phi)$, $q_1^\star(\Phi)$, $q_2^\star(\Phi)$, and equilibrium profits, $\Pi_1^\star(\Phi)$ and $\Pi_2^\star(\Phi)$.

Data-sharing stage. Recall that data sharing by firm 2 generates a greater data advantage in market A. Intuitively, this allows firm 1 to collect more data in market A which can then be leveraged to enhance its data advantage in market B to the detriment of firm 2. Hence, for firm 2, sharing its market B data with firm 1 would seem to be an act of self-sabotage. Contrary to this wisdom, we find the following result.

Proposition 1 (**Profitable Data Sharing**). *After the entry of firm* 1 *in market* B*, firm* 2 *is willing to share its data with firm* 1*, even for free:* $\Pi_2^\star(1) > \Pi_2^\star(0)$.

Proposition 1 is a novel result in the literature. Firm 2 achieves higher profits by sharing data with firm 1. With this action, firm 2 transforms firm 1 from a *competitor* into a *co-opetitor*. Firm 2's increased data collection in market B benefits 1 in A, where it is a monopolist.[22] To unpack the intuition behind this result, we examine the impact of firm 2 sharing its data on equilibrium quality investment and output choices.

Corollary 1 (**Quality Improvements, Output and Data Sharing**). *The receipt of data from market* B *induces firm* 1 *to scale back its operations in market* B *and firm* 2 *to expand:*

$$v_1^\star(1) < v_1^\star(0),\ q_1^\star(1) < q_1^\star(0) \quad \textit{and} \quad v_2^\star(1) > v_2^\star(0),\ q_2^\star(1) > q_2^\star(0). \tag{9}$$

In market A*, the receipt of data from market* B *induces firm* 1 *to scale back its operations if cross-market externalities are sufficiently strong:*

$$v_A^\star(1) < v_A^\star(0) \Leftrightarrow \theta > \theta^S \approx 0.27 \quad \textit{and} \quad q_A^\star(1) < q_A^\star(0) \Leftrightarrow \theta > \theta^Q \approx 0.31. \tag{10}$$

When firm 2 shares data the *data-sharing effect* reduces firm 1's incentive to increase quality improvement in both markets because the data it receives from firm 2 substitutes for its own costly investment in quality improvements. Firm 2 optimally responds by increasing its investment (and output) in market B. The effects of data sharing on output largely follow the effects on quality improvement in market B. Indeed, as data sharing lowers firm 1's quality improvement efforts in B, its output in market B also falls. Because output choices in B are strategic substitutes, firm 2 increases production, which is further enhanced by the increased investment by firm 2 in market B. Together, the changes in investment and output in market

[22] This mechanism may explain why the European Commission (2018) observed that firms share data freely with other firms (see pp.60-65 for a summary of data sharing practices).

B show how firm 1 *accommodates* its data-sharing rival. Hence, data sharing is a way for firm 2 to soften the aggressiveness of its new competitor.

In market A, firm 1 faces fewer forces on its incentives to invest in quality improvements. The overall effect of data sharing depends on the rate at which data translates into increased consumers' willingness to pay, θ. When sufficiently low, $\theta < \theta^S \approx 0.27$, the *data-sharing effect* is relatively weak and firm 1 improves the quality of its product in market A, compensating for a reduction in quality improvements in market B. But when high, $\theta > \theta^S \approx 0.27$, the *data-sharing effect* dominates and firm 1 invests less in quality improvements even in market A (and hence in both markets) when firm 2 shares its data. The results for output in market A qualitatively follow the effects on quality improvements levels in market A. Specifically, when the rate at which data translates into increased consumers' willingness to pay (θ) is high, in addition to lowering output in market B, firm 1 further commits to lower data advantages to its affiliate in market B. This willingness of firm 1 to lower its quality improvement efforts in A as well as its data advantage in market B is a particularly acute manifestation of its incentive to soften competitive forces on firm 2 in lieu of higher volume of data. As a consequence of this, firm 2 prefers to share its data.

Welfare effects of data sharing. It follows from Proposition 1 that industry profits are higher when firm 2 shares its data. The effect on consumers is less immediate. Consumers in a given market are better off when the total output in that market rises. Therefore, Corollary 1 tells us consumer surplus in market A falls when externalities are strong ($\theta > \theta^Q$), and rises when they are not. Corollary 1 also tells us that when firm 2 shares data, firm 1 produces less in market B, while firm 2 produces more. Corollary 2 summarizes.

Corollary 2 (Industry Profit and Consumer Surplus). *When firm 2 shares its data:*

1. *industry profit is higher;*
2. *consumer surplus is: (i) lower in market A if and only if $\theta > \theta^Q$; (ii) lower in market B.*

The Impact of Entry by Firm 1 in Market B. Comparing the benchmark case to our main analysis (Sections 4.1 and 4.2), we can see how entry by firm 1 in market B impacts the data sharing decision of firm 2. Doing so, we uncover a novel rationale for market entry: entering market B (and becoming a generalist) firm 1 prompts firm 2 to share data on more favorable terms. After firm 1's entry in market B, firm 2 is no longer indifferent and instead gains from sharing data with 1, even for free. Combining these observations gives Proposition 2.

Proposition 2 (Entry Rationale). *Entry by firm 1 increases firm 2's incentive to share data.*

By being active in market B and exerting competitive pressure on firm 2, firm 1 prompts firm 2 to share its data because it softens competition in market B. To underscore the nature of this result, consider a case in which firm 1 has full bargaining power. Firm 2 would *pay* an amount equal to $\Pi_2^\star(1) - \Pi_2^\star(0) > 0$ to share its data with firm 1.

5. Model extensions

In this section we present several extensions that probe the robustness of our model. We consider each of the following: (i) price competition; (ii) same-side network effects; (iii) heterogeneous cross-market externalities; and (iv) two generalist firms. Details, proofs and calculations are in our online-only Appendix B.[23]

5.1. Differentiated Bertrand competition

Our main model features classic Cournot competition so that products are perfect substitutes. Our insights about data sharing also hold for typically differentiated products. We show this with a demand system arising from the well-known Hotelling set-up in which firms choose prices. There is a unit mass of consumers in each market. In market A, their valuations are distributed according to an outside option, $s \sim \mathcal{U}[0,1]$.

$$u_A(v_A, p_A, \Phi, s) = v_A + \theta(q_1^e + \Phi q_2^e) - p_A - s, \tag{11}$$

[23] In addition, in our (also online-only) Appendix C we combine extensions (i) and (ii) by solving the model with both price competition and same-side network effects. In that appendix, we also show that our main result is robust to the case when allowing for consumer expectations, à la Katz & Shapiro (1985).

where p_A is the price charged by the generalist in market A, Φ is the indicator function whether the specialist firm 2 in market B shares data and q_1^e and q_2^e is the expected value consumers get from data collected in market B. Consumers buy when $u_A(\cdot) \geq 0 \implies s < \tilde{s}(v_A, p_A, \Phi)$, which determines demand in market A.[24] In market B, the generalist firm 1 and the incumbent specialist firm 2 compete to attract consumers who are distributed uniformly on a unit-length Hotelling preference line.[25] Firm 1 is located at 0 and firm 2 is located at 1.

The utility of a consumer of type x from firm 1 and 2's products are, respectively,

$$u_1(v_1, p_1, x) = v_1 + \theta q_A^e - p_1 - tx, \qquad u_2(v_2, p_2, x) = v_2 - p_2 - t(1 - x), \tag{12}$$

where p_i and v_i are the price and the value from investments at firm $i \in \{1,\ 2\}$, t is the transportation cost parameter and q_A^e is the expected value from data collected in market A. Consumers demands are constructed by identifying the consumer that is indifferent between buying from firm 1 or firm 2 and is denoted as $\tilde{x}$, i.e., $u_1(\cdot) = u_2(\cdot) \implies \tilde{x}(v_1, v_2, p_1, p_2, q_A^e)$. Thus, in this framework we can represent the demands as

$$q_1(v_1, v_2, p_1, p_2, q_A^e) = \tilde{x}(\cdot), \quad q_2(v_2, v_1, p_2, p_1, q_A^e) = 1 - \tilde{x}(\cdot). \tag{13}$$

Note that consumer demands are now directly affected by the value generated from data collected in market A. Specifically, as the data collected in market A increases, the demand for firm 1's product increases while the demand firm 2's product decreases. This direct effect of data on demand for the product of the two firms is a novel feature of this demand system which was absent in our main, Cournot, setting.

The profits of the firms are calculated as before:

$$\Pi_1 = \underbrace{p_A q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1 q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2 q_2(\cdot) - I(v_2). \tag{14}$$

[24]To ensure that demand in market A is elastic, we assume (as specified precisely by Assumption 1) that this market is not covered such that some consumers do not purchase the product in equilibrium.

[25]For firms to compete in market B, we assume (as specified precisely by Assumption 1) that market is covered.

The timing of the game is as follows: (i) Firms set investments v_1, v_A and v_2; (ii) Firms set prices p_1, p_A, p_2; (iii) Consumers form expectations on the value generated by data then decide whether to buy.[26] We impose the following technical restrictions.

Assumption 1. *(i) The value of cross-network data externality θ is sufficiently low, i.e., $\theta < \tilde{\theta} \approx 0.322$; (ii) the transportation cost parameter is within the region $\frac{2+15\theta^2}{18} < t < \frac{8-9\theta^2+\sqrt{64+225\theta^4-144\theta^2}}{72}$.*

These restrictions ensure that the second-order conditions are satisfied, and that we get an interior solution. In particular, under these conditions, market A is not covered so that the demand is elastic and the generalist firm invests in innovation and benefits from the data of the specialist firm. On the contrary, market B is covered and firms compete in this market. We present a detailed analysis in Appendix B, showing that firm 2 prefers to share data with firm 1, $\Pi_2^\star(1) - \Pi_2^\star(0) > 0$, as per our main result.

5.2. Same-side data externalities

The mechanism we reveal with our main model is a consequence of cross-market externalities. Same-side network effects may also exist, where a product improves when there is more consumer data generated from a product's own use. In this section, we show that our main result is qualitatively unaffected with the addition of same-side network effects. Consider the following inverse demand functions

$$P_A(v_A, q_1, \Phi q_2, q_A) = \mathcal{A} + v_A + \theta(q_1 + \Phi q_2) - \beta_A q_A, \tag{15}$$

$$P_1(v_1, q_A, q_1, Q_B) = \mathcal{B} + v_1 + \theta q_A + \sigma q_1 - \beta_B Q_B, \tag{16}$$

$$P_2(v_2, q_2, Q_B) = \mathcal{B} + v_2 + \sigma q_2 - \beta_B Q_B, \tag{17}$$

where $\sigma \geq 0$ represents the strength of the same-side externality in market B. Letting $\sigma = 0$ nests our main specification. Adding these terms comes at a considerable cost to brevity.

[26] Expectations are assumed to be correct in equilibrium, as usual. To address out-of-equilibrium cases, we assume expectations are "responsive" in the sense that they are correct for any choices of firms at prior stages.

Nevertheless, and as detailed in Appendix B, we cover and here report two tractable cases.

The first is a "neutral" or "balanced" case, in which data from market A are equally helpful as those from B for product enhancement, i.e., $\sigma = \theta$. In a second natural case of interest, we consider data from the same market to be more helpful for product development, $\sigma > \theta$. In both cases, our main result prevails: firm 2 strictly prefers to share its data.

The presence of an intra-market externality in market B adds an incentive for firms to produce more. All else equal, a firm producing more would typically lower the market price (and this still happens in our model because σ must be sufficiently small to produce well-defined solutions) but the product enhancements raise consumers' willingness to pay, which encourages the firm to set higher prices. Overall, this makes price less sensitive to output, and so output in market B increases in σ. This increases the relative importance of a firm's own production in market B in their total profits.

This does not overturn the willingness of firm 2 to share its data, because doing so still prompts firm 1 to accommodate firm 2 in B. In particular, the effects described by (9) persist. As such, the mechanism we uncover is present in the face of the intra-market externality. However, an increase in its strength (σ) reduces the extent to which firm 1 accommodates firm 2, and as such, firm 2's profit falls with σ. Overall, the intra-market externality reduces the potency of the forces we identify with our main result, but does not overturn them.

5.3. Heterogeneous cross-market externalities

Our contribution relies crucially on cross-market externalities, and our important parameter there is $\theta > 0$: the rate at which data from sales of one product translates into a higher willingness to pay for another product. We assumed that the rates at which data from market A affects product B, and vice versa, to be equal. This is inaccurate when data from A are more useful in improving product B than data from B are in improving product A, or vice versa; or, when the rate at which a company can take advantage of data is different for data sourced internally (as is the case for firm 1) versus externally (when firm 2 shares with firm 1).[27]

[27]This is the case, e.g., in Baldwin (2018), who models a one-way externality.

To address this, here we consider asymmetric cross-market externalities. Namely, data collected in market A and shared in B induce an externality proportional to θ_A, while data collected in B and shared in A induce an externality proportional to θ_B, which gives demands:

$$P_A(v_A, q_1, \Phi q_2, q_A) = \mathcal{A} + v_A + \theta_B(q_1 + \Phi q_2) - \beta_A q_A, \tag{18}$$

$$P_1(v_1, q_A, q_1, Q_B) = \mathcal{B} + v_1 + \theta_A q_A - \beta_B Q_B, \tag{19}$$

$$P_2(v_2, q_2, Q_B) = \mathcal{B} + v_2 - \beta_B Q_B. \tag{20}$$

As we detail in Appendix B, different cross-market externalities do not change our model's qualitative results. What matters for the mechanism we reveal is that data from one market improves quality in the other, i.e., that $\theta_A, \theta_B > 0$, not the relative rates at which they do so.

5.4. Two competing generalist firms

In the baseline, we demonstrated that our main results in a setting in which the specialist firm faces one generalist competitor. We now show that the core effect at play when firm 2 decides to share its data remains unchanged when markets A and B have multiple (2) generalist firms.

The two generalist firms are indexed as firms 1 and 3, while the specialist remains indexed as firm 2. Indicators Φ_1 and Φ_3 capture the decision of firm 2 to share data with firm 1 and firm 3, respectively (a value of 1 indicates data are shared). The total outputs are now $Q_A = q_{A1} + q_{A2}$ in market A and $Q_B = q_{B1} + q_2 + q_{B3}$. The inverse demand functions in markets A and B are:

$$P_{A1}(v_{A1}, q_{B1}, \Phi_1 q_2, Q_A) = \mathcal{A} + v_{A1} + \theta(q_{B1} + \Phi_1 q_2) - \beta_A Q_A, \tag{21}$$

$$P_{B1}(v_{B1}, q_{A1}, Q_B) = \mathcal{B} + v_{B1} + \theta q_{A1} - \beta_B Q_B, \tag{22}$$

$$P_2(v_2, q_2, Q_B) = \mathcal{B} + v_2 - \beta_B Q_B, \tag{23}$$

$$P_{A3}(v_{A3}, q_{B3}, \Phi_3 q_2, Q_A) = \mathcal{A} + v_{A3} + \theta(q_{B3} + \Phi_3 q_2) - \beta_A Q_A, \tag{24}$$

$$P_{B3}(v_{B3}, q_{A3}, Q_B) = \mathcal{B} + v_{B3} + \theta q_{A3} - \beta_B Q_B. \tag{25}$$

As usual, the resulting profits of the firms can be written:

$$\Pi_1 = P_{A1}(\cdot)q_{A1} + P_{B1}(\cdot)q_{B1} - I(v_{A1}) - I(v_{B1}), \tag{26}$$

$$\Pi_2 = P_2(\cdot)q_2 - I(v_2), \tag{27}$$

$$\Pi_3 = P_{A3}(\cdot)q_{A3} + P_{B3}(\cdot)q_{B3} - I(v_{A3}) - I(v_{B3}). \tag{28}$$

We focus our discussion on the decision of firm 2 to share its data, and on the willingness of the generalist firms to accept the data. Denote by $\Pi_1^*(\Phi_1, \Phi_3)$, $\Pi_2^*(\Phi_1, \Phi_3)$, and $\Pi_3^*(\Phi_1, \Phi_3)$ the profits of firms 1, 2, and 3 depending on whether firm 2 shares its data with firm 1 and firm 3.

We prove that the specialist firm shares its data with both generalists. To do so, we first confirm that firm 2 is willing to share its data with both competitors:

$$\Pi_2^*(1,1) > \Pi_2^*(0,1) = \Pi_2^*(1,0) > \Pi_2^*(0,0). \tag{29}$$

Note that sharing data with one of the firms, and not the other, is less profitable for firm 2 than sharing data with both generalist firms. It remains to show that both generalist firms are willing to accept the data. Compared to a situation in which neither generalist firm has data from firm 2, it is profitable for a generalist to accept the data:[28]

$$\Pi_1^*(1,0) > \Pi_1^*(0,0) \text{ and } \Pi_3^*(0,1) > \Pi_3^*(0,0). \tag{30}$$

The effects at play are as in our main analysis. Sharing incentivizes generalist firms to accommodate the specialist by lowering their output in market B. This prompts firm 2 to expand and share more data, and the fact firms compete in market A does not remove this effect.

[28]Also, when one generalist receives data from firm 2, the other benefits from those data too: $\Pi_1^*(1,1) > \Pi_1^*(0,1)$ and $\Pi_3^*(1,1) > \Pi_3^*(1,0)$.

5.5. Long-term competitive dynamics.

The entry of the generalist firm into a secondary market may be followed by market changes that alter the firm's objectives. If market B becomes larger or more lucrative over time, firm 1's objectives might well shift to full domination of market B and threatening the survival of firm 2. Our analysis thus far assumes a one-shot game featuring data sharing and competition, and in which data-sharing is shown to be a source of co-opetition. One topic for further research is to consider how the forces we identify interact with features relevant to a more dynamic competitive setting. In the relevant markets, one particularly plausible outcome when each firm is uncertain about its future production and innovation costs, is the exit of specialist firms. Consumer data are typically used to better direct their innovation and understand future trends of the markets. Hence, access to large amounts of consumer data can allow a firm to reduce uncertainty and lower the chance of exit, while increasing its own dominance and potentially increasing rivals' rate of exit. In this case, would a specialist firm still be willing to share its data if its larger rival could use it in the long run to cement its dominance and force the specialist out of business?

While a full analysis of the long run is outside the scope of this article, we discuss some of the different factors at play in light of our results. On the one hand, if data allow a firm to lower its production costs over time, a small firm could be reluctant to share its data and increase the chances of being pushed out of the market by an increasingly efficient rival. On the other hand, we showed that a specialist firm innovates more when it shares its data, while the generalist rival invests less. These shifts in investment may have long-lasting positive impacts on the profits of the specialist firm. Hence, the willingness of a specialist to share data when it accounts for such competitive dynamics crucially depends on the respective importance of data and investments in innovation on future production costs and expected profits.

A second and related factor is acquisitions. If data-sharing increases a rival's competence or efficiency in a specialist firm's market, then it may increase the likelihood of a takeover. Some smaller or specialist companies seem, from their inception, to wish to be acquired. For companies with such shorter-term goals, it is not clear if the mechanism we uncover is beneficial.

While data-sharing may eventually lead to a lower-priced acquisition (or an exit), the increased flow of profit during a preceding phase of co-opetition could compensate for this. For specialist companies with longer-term objectives the calculus seems yet more complicated as more factors become relevant, but for them data-sharing may be distinctly less attractive to the extent that a limited period of co-opetition is not a strong enough reason to share data.

With high rates of acquisitions and such dominant acquirers, many of the markets relevant to this article are also a natural focus of competition authorities. Our results suggest that data-sharing, by reducing competitive pressures and aligning incentives between rival firms, could delay or reduce acquisitions, perhaps pushing towards less concentrated markets. It is a stretch to suggest that data-sharing is equivalent to ownership, but our analysis highlights that if data-sharing leads to more firms in a market, this should not be interpreted as unambiguously pro-competitive and could in fact be a symptom of co-opetition.

Overall, whether data sharing is still more profitable than data hoarding when accounting for long-term competitive dynamics would seem to depend on the type of markets considered, and the horizons and objectives of the firms involved.

6. Implications and discussions

Managerial implications. In general, our work highlights the importance for data-driven companies to put data governance considerations at the center of their business models, and to acknowledge the strategic role of data.

Managerial Insight 1. *When a large firm enters a new market to benefit from cross-market data exernalities, it may be profitable for a smaller incumbent to share its data with it, even for free. Data sharing can be a strategic device to lower the multi-market firm's aggressiveness.*

With this point, we offer stark and a priori counter-intuitive advice to managers of specialist firms. Sharing value-enhancing data with a multi-market rival can in fact afford some breathing space, potentially much needed in today's competitive digital landscape. Specifically, our results show that data-sharing can make even a formidable rival a less aggressive competitor because it makes them internalize the value of the data shared with them.

Turning to generalist firms, our advice may also run counter to first instincts:

Managerial Insight 2. *A generalist (multi-market) firm can increase its overall profit by giving up market share to a specialist (single-market) rival, when the specialist shares its data.*

Giving up market share is surely bad standalone advice, but our work shows that doing so in one market can be more than compensated by increased profits in another. When a specialist is willing to share their data, it may be profitable to respond by accommodating the rival because it increases the flow of data that can be used to increase the value proposition to customers. The increased data collection substitutes for more costly investments, improving bottom lines.

Last, we propose a new consideration for firms considering entry into new markets:

Managerial Insight 3. *Entry into a market for which data are related to the firm's existing operations can result in more favorable data-sharing agreements with incumbent specialist firms.*

This suggests that firms should explore opportunities in data-relevant markets. In addition to collecting relevant data, their presence in the new market also gives them the leverage to negotiate more favorable terms for data sharing with incumbent specialist firms. This shows a new source of value to establishing a subsidiary in data-relevant markets, and so constitutes a new strategic rationale for entry in the presence of cross-market externalities.

Policy implications and regulation. Recognizing the importance of data as a source of competitive advantage, there is a growing policy discussion and academic literature on optimal regulation in the presence of data-driven network effects (Crémer et al. 2019, Hagiu & Wright 2020, Krämer & Schnurr 2022, Krämer & Shekhar 2022, Prüfer & Schottmüller 2021, Parker et al. 2021, Tucker 2019).

We predict that by sharing its data, the specialist firm boosts its sales and profits due to the strategic retreat of the generalist firm in that market. In general terms, we show data sharing to be a form of "puppy-dog" strategy by specialists that softens the aggressiveness of "fat cat" generalists (in the spirit of Fudenberg & Tirole 1984).

Both firms in our model profit from data sharing. To competition authorities, information exchange between firms is a classic signal of collusion (Awaya & Krishna 2016, Cason & Mason 1999, Clarke 1983, Kandori & Matsushima 1998). But in our setting, data sharing lowers competition without any explicit underlying collusion.

Data-sharing can unleash economies of scope and provide benefits across markets—a message well-received by policy makers. For example, the European Commission (2022) reports:

"As data is a non-rivalrous resource, it is possible for the same data to support the creation of several new products, services or methods of production. So, companies can engage with the same data in different arrangements with other big companies, small and medium-sized enterprises (SMEs), startups or the public sector. This way, the value resulting from the data can be fully exploited."

Those benefits seem undeniable (and are captured by our model's externalities). In line with that general and positive sentiment, in our analysis firms benefit from sharing and the increased data availability leads to improved products and services. Those improvements increase the total surplus available and may yield further beneficial effects outside the scope of our study.

Yet, the benefits will remain unrealized if firms are resistant to data sharing. With a view to unlocking the gains from sharing, reports such as that by the European Commission (2020*a*) emphasize when it may harm the sharer and constitute a market failure.[29] To those points, our analysis shows data sharing to be an important strategic decision for firms that can be mutually profitable for both sharer and receiver, even when they have asymmetric market positions:

Policy Insight 1. *Data sharing can reduce market concentration and enhance the market share of smaller specialist firms.*

Data sharing is a double-edged sword for policy. On the one hand, it encourages generalist firms to accommodate specialists, which increases their investments in innovation. On the other hand, it lowers consumer surplus as investment intensity drops and firms fall into a *co-opetitive*

[29]Related to our work, that report points out an incentive for firms to share data across markets when goods are complements (European Commission 2020*a*, pp.20-21).

relationship. This presents a dilemma for policy and data-sharing policies should acknowledge that increased market share of specialist firms in market B does not imply fiercer competition.

Policy Insight 2. *When a specialist (single-market) firm shares its data with a competitor that operates in two markets, consumer welfare falls in the market in which they compete. Consumer welfare can also fall in the primary market of the generalist.*

Finally, our results suggest that reduced competition caused by data-sharing also has long-term implications. This follows from the lower innovation intensities data sharing prompts.

Policy Insight 3. *Data sharing may lower innovation in both markets when the value of inter-market data is high.*

The negative impacts of data sharing we reveal call for caution over recent regulations such as the European Commission's Data Act[30] in which Chapter II standardizes data access between firms, and the proposed Data Governance Act,[31] which aims to facilitate B2B data exchanges.

Another issue intimately related to data sharing and often near the top of the list of concerns is that of consumer privacy. For example, limits on information-sharing practices are included in the California Consumer Privacy Act[32] and the European General Data Protection Regulation.[33] That said, technological advances are providing an increasing number of privacy-preserving data-sharing solutions, which may reduce the tension between privacy and data sharing.[34] When firms cannot share data, our work suggests that consumers may in fact benefit, not (only) from privacy protection, but from more intense competition between firms.

In sum, our insights present tensions for policy. It is not within the scope of our analysis to provide a complete resolution to these tensions, especially as these markets and the associated technologies continue to develop rapidly. But our analysis does show that any blanket support for B2B data sharing would seem misguided. More generally, we hope to bring the competitive consequences of cross-market data-sharing agreements to the forefront of the debate.

[30] https://digital-strategy.ec.europa.eu/en/policies/data-act, last accessed November 7, 2022.
[31] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020PC0767, last accessed November 7, 2022.
[32] https://theccpa.org/, last accessed Sep 27, 2022.
[33] https://gdpr.eu/, last accessed Sep 27, 2022.
[34] See, for example, the Deloitte Insights 2022 report here: https://www2.deloitte.com/xe/en/insights/focus/tech-trends/2022/data-sharing-technologies.html, last accessed Jan 19, 2023.

7. Concluding discussion

Business-to-business (B2B) data sharing is increasingly common. In contrast to oft discussed advantages like information synergies and value creation, we unveil a new strategic rationale for B2B data sharing. We show that specialist (single-market) firms can employ data-sharing commitments to mitigate competition by transforming their multi-market competitors into cooperative partners (co-opetitors). This strategy allows small incumbent firms to better cope with the entry of large, multi-market competitors in their market.

However, we caution that the approach may not always yield positive outcomes, and managers need to discern the conditions under which such data sharing is profitable. Specifically, if data sharing fails to diminish the aggressive nature of a multi-market rival, then the smaller firm may find sharing data unprofitable. This is particularly true in scenarios where the technology is already mature, and there is limited potential for medium-term quality improvements. In such cases, smaller firms may refrain from data sharing, as their investments would have negligible impact on reducing the aggressiveness of the dominant firm in the market they compete in.

We hope this work can be used and built on to further explore how data sharing affects market outcomes. One direction is to consider richer competitive dynamics. For instance, consider the incentives of multi-market firms to acquire specialist competitors (especially relevant given how many acquisitions occur in digital markets). On the one hand, data sharing may help firms better anticipate the value of information synergies resulting from a merger (see, e.g., Dubus & Legros 2022) and increase the overall value of mergers and acquisitions. On the other hand, after data are shared, the benefits of a merger are reduced compared to the no-sharing case in which firms compete head-to-head. Once data are shared, as we have shown, firms may compete less aggressively. This lowers the competition-reducing gains of a merger in the case data were already shared before the merger (vis-á-vis no sharing before the merger). This suggests that after data are shared, firms are more likely to merge for efficiency gains than for a weaker anti-competitive effect, an issue relevant to firms and regulators alike.

Appendix A. Proofs for the main model

In this appendix we prove the results of our main model, introduced in Section 3. In doing so, we also provide the exact expressions behind some of the terms we referenced in the main text.

Proof of Lemma 1. Differentiating the profit of firm 1 with respect to q_A and q_1 and the profit of firm 2 with respect to q_2 yields the following system of first-order conditions:

$$\frac{\partial \Pi_1}{\partial q_k} = \underbrace{P_k(\cdot)}_{\text{Volume effect}} + \underbrace{\frac{\partial P_k(\cdot)}{\partial q_k} q_k}_{\text{Margin effect}} + \underbrace{\frac{\partial P_j(\cdot)}{\partial q_k} q_j}_{\substack{\text{Value increase in market B} \\ \text{from market A data}(+)}} = 0, \text{ for } k \neq j \in \{A, 1\}, \tag{A.1}$$

$$\frac{\partial \Pi_2}{\partial q_2} = \underbrace{P_2(\cdot)}_{\text{Volume effect}} + \underbrace{\frac{\partial P_2(\cdot)}{\partial q_2} q_2}_{\text{Margin effect}} = 0. \tag{A.2}$$

Solving yields the equilibrium output choices as functions of sharing and investment decisions:

$$\hat{q}_A(v_A, v_1, v_2, \Phi) = [2\omega]^{-1}\Big[6(\mathcal{A} + v_A) + \theta(\mathcal{A}(2 + \Phi) + (4 - \Phi)v_1 - 2(1 - \Phi)v_2)\Big], \tag{A.3}$$

$$\hat{q}_1(v_1, v_A, v_2, \Phi) = [2\omega]^{-1}\Big[4\mathcal{A} + 8v_1 - 4v_2 + 4\theta(\mathcal{A} + v_A) + \Phi\theta^2(\mathcal{A} + v_2)\Big], \tag{A.4}$$

$$\hat{q}_2(v_2, v_1, v_A, \Phi) = [\omega]^{-1}\Big[\mathcal{A}(2 - \theta - \theta^2) - 2v_1 - \theta v_A + v_2(4 - \theta^2)\Big], \tag{A.5}$$

$$\begin{aligned} \hat{Q}_B(v_2, v_1, v_A, \Phi) &= \hat{q}_1(\cdot) + \hat{q}_2(\cdot) \\ &= [2\omega]^{-1}\Big[4(2\mathcal{A} + v_1 + v_2) + 2\theta(\mathcal{A} + v_A) - (2 - \Phi)\theta^2(\mathcal{A} + v_2)\Big], \end{aligned} \tag{A.6}$$

$$\omega \equiv 12 - (4 - \Phi)\theta^2. \tag{A.7}$$

Because investment creates value (shifts demand out), firms produce more in a given market if they invest more in their product in that market:

$$\frac{\partial \hat{q}_1(\cdot)}{\partial v_1} = \frac{4}{\omega} > 0, \ \frac{\partial \hat{q}_A(\cdot)}{\partial v_A} = \frac{3}{\omega} > 0, \ \frac{\partial \hat{q}_2(\cdot)}{\partial v_2} = \frac{4 - \theta^2}{\omega} > 0. \tag{A.8}$$

Due to the cross-market data advantage, if firm 1 invests more in one market and therefore produces more in that market, its product in the other market also improves, boosting demand. Therefore, firm 1 increases output there too:

$$\frac{\partial \hat{q}_A(\cdot)}{\partial v_1} = \frac{(4-\Phi)\theta}{2\omega} > 0, \;\; \frac{\partial \hat{q}_1(\cdot)}{\partial v_A} = \frac{2\theta}{\omega} > 0. \tag{A.9}$$

In contrast, firms produce less when their rival invests more in quality improvement. For example, if firm 1 invests more in market B, it produces more in B. This increases the competitive pressure on firm 2, and as output choices are strategic substitutes, it produces less. In sum:

$$\frac{\partial \hat{q}_A(\cdot)}{\partial v_2} = \frac{\theta(\Phi-1)}{\omega} \leq 0, \frac{\partial \hat{q}_1(\cdot)}{\partial v_2} = \frac{\Phi\theta^2-4}{2\omega} < 0, \frac{\partial \hat{q}_2(\cdot)}{\partial v_A} = -\frac{\theta}{\omega} < 0, \;\; \frac{\partial \hat{q}_2(\cdot)}{\partial v_1} = -\frac{2}{\omega} < 0. \tag{A.10}$$

Substituting these outputs into demand, we can then write the firms' profits as:

$$\hat{P}_A(v_A, v_1, v_2, \Phi) = P_A(v_A, \hat{q}_1(\cdot), \Phi\hat{q}_2(\cdot), \hat{q}_A(\cdot)), \tag{A.11}$$

$$\hat{P}_1(v_A, v_1, v_2, \Phi) = P_1(v_1, \hat{q}_A(\cdot), \hat{Q}_B(\cdot)), \tag{A.12}$$

$$\hat{P}_2(v_2, v_1, v_A, \Phi) = P_1(v_2, \hat{Q}_B(\cdot)). \tag{A.13}$$

$$\max_{v_A, v_1} \hat{\Pi}_1(v_A, v_1, v_2, \Phi) = \hat{P}_A(\cdot)\hat{q}_A(\cdot) + \hat{P}_1(\cdot)\hat{q}_1(\cdot) - I(v_A) - I(v_1) \tag{A.14}$$

$$\max_{v_2} \hat{\Pi}_2(v_2, v_1, v_A, \Phi) = \hat{P}_2(\cdot)\hat{q}_2(\cdot) - I(v_2). \tag{A.15}$$

Now we demonstrate that keeping investments constant, firm 2 has no incentive to share data with the generalist firm 1, $\hat{\Pi}_2(\cdot, \Phi = 1) < \hat{\Pi}_1(\cdot, \Phi = 0)$. In particular, we find

$$\frac{\partial \hat{\Pi}_2(\cdot, \Phi)}{\partial \Phi} = -\frac{4\theta^2(\mathcal{A}(2-\theta-\theta^2) + v_2(4-\theta^2) - 2v_1 - \theta v_A)^2}{(12-\theta^2(4-\Phi))^2} < 0, \tag{A.16}$$

which establishes that $\hat{\Pi}_2(\cdot, \Phi = 1) < \hat{\Pi}_1(\cdot, \Phi = 0)$. ■

Proof of Proposition 1. The profits of the firms are given by (A.14) and (A.15). Applying the envelope theorem to the first-order conditions, yields the following, for $k \in \{A, 1\}$:

$$\frac{\partial \hat{\Pi}_1(\cdot)}{\partial v_k} = \underbrace{\underbrace{\frac{\partial P_k(\cdot)}{\partial v_k}\hat{q}_k(\cdot)}_{\text{Margin effect}} - \underbrace{\frac{\partial I(v_k)}{\partial v_k}}_{\text{Cost}}}_{\text{Direct effects}} + \underbrace{\underbrace{\frac{\partial P_1(\cdot)}{\partial Q_B}\frac{\partial \hat{q}_2(\cdot)}{\partial v_k}\hat{q}_1(\cdot)}_{\text{Competitive effect } (+)} + \underbrace{\overbrace{\frac{\partial P_A(\cdot)}{\partial q_2}}^{=\Phi}\frac{\partial \hat{q}_2(\cdot)}{\partial v_k}\hat{q}_A(\cdot)}_{\text{Data-sharing effect } (-)}}_{\text{Strategic effects (?)}} = 0, \tag{A.17}$$

$$\frac{\partial \hat{\Pi}_2(\cdot)}{\partial v_2} = \underbrace{\underbrace{\frac{\partial P_2(\cdot)}{\partial v_2}\hat{q}_2(\cdot)}_{\text{Margin effect}} - \underbrace{\frac{\partial I(v_2)}{\partial v_2}}_{\text{Cost}}}_{\text{Direct effects}} + \underbrace{\frac{\partial P_2(\cdot)}{\partial Q_B}\frac{\partial \hat{q}_1(\cdot)}{\partial v_2}\hat{q}_2(\cdot)}_{\text{Competitive effect } (+)} = 0. \tag{A.18}$$

Solving gives the equilibrium quality improvement levels as functions of sharing decision Φ:

$$\begin{aligned} v_A^\star(\Phi) = \Omega^{-1}\Big[&2\mathcal{A}(56 + \Phi^2\theta^3 + 2\Phi\theta(2-\theta)(3+\theta(4+\theta(3+\theta))) \\ &+ 4\theta(8-\theta(8-\theta(6-\theta-\theta^2))))\Big], \end{aligned} \tag{A.19}$$

$$v_1^\star(\Phi) = \Big[2\Omega\Big]^{-1}\Big[\mathcal{A}(192 + \theta\mathcal{H})\Big], \tag{A.20}$$

$$v_2^\star(\Phi) = \Omega^{-1}\Big[4\mathcal{A}(4-\theta^2)(6-\theta(12+4\theta(4-\theta-\theta^2)) + \Phi(1+\theta(2-\theta-\theta^2)))\Big], \tag{A.21}$$

$$\begin{aligned} \mathcal{H} \equiv\ &512 - 80g - 24\theta(6-\Phi(10-\Phi)) - 4\theta^2(4-\Phi)(20-\Phi) \\ &+ 2\theta^3(12-\Phi(62-\Phi(18-\Phi))) + 4\theta^4(4-\Phi)(3-\Phi) + \Phi\theta^5(4-\Phi)^2, \end{aligned} \tag{A.22}$$

$$\begin{aligned} \Omega \equiv\ &336 - 4\theta^2(176-\Phi(71-\Phi)) + \theta^4(344-\Phi(236-\Phi(40-\Phi))) \\ &- (6-\Phi)(4-\Phi)(2-\Phi)\theta^6. \end{aligned} \tag{A.23}$$

Substituting into outputs and profits yields

$$q_A^\star(\Phi) = \hat{q}_A(v_A^\star(\Phi), v_1^\star(\Phi), v_2^\star(\Phi), \Phi), \tag{A.24}$$

$$q_1^\star(\Phi) = \hat{q}_1(v_1^\star(\Phi), v_A^\star(\Phi), v_2^\star(\Phi), \Phi), \tag{A.25}$$

$$q_2^\star(\Phi) = \hat{q}_2(v_2^\star(\Phi), v_1^\star(\Phi), v_A^\star(\Phi), \Phi), \tag{A.26}$$

$$Q_B^\star = q_1^\star(\Phi) + q_2^\star(\Phi). \tag{A.27}$$

And substituting these equilibrium outputs into the inverse demand expressions gives

$$P_A^\star(\Phi) = \hat{P}_A(v_A^\star(\Phi), v_1^\star(\Phi), v_2^\star(\Phi), \Phi), \tag{A.28}$$

$$P_1^\star(\Phi) = \hat{P}_1(v_1^\star(\Phi), v_A^\star(\Phi), v_2^\star(\Phi), \Phi), \tag{A.29}$$

$$P_2^\star(\Phi) = \hat{P}_2(v_2^\star(\Phi), v_1^\star(\Phi), v_A^\star(\Phi), \Phi). \tag{A.30}$$

The equilibrium profit of firm 1 and 2 is

$$\Pi_1^\star(\Phi) = P_1^\star(\Phi)q_1^\star(\Phi) + P_A^\star(\Phi)q_A^\star(\Phi) - I(v_1^\star(\Phi)) - I(v_A^\star(\Phi)), \tag{A.31}$$

$$\Pi_2^\star(\Phi) = P_2^\star(\Phi)q_2^\star(\Phi) - I(v_2^\star(\Phi)). \tag{A.32}$$

Comparing the profit of firm 2 when it does versus does not share its data,

$$\Pi_2^\star(1) - \Pi_2^\star(0) = \frac{\mathcal{A}^2\theta\mathcal{G}}{18\mathcal{T}^2}, \tag{A.33}$$

$$\begin{aligned}\mathcal{G} \equiv{}& 423360 - 16400016\theta + 18224664\theta^2 + 46692120\theta^3 - 180178080\theta^4 - 108524320\theta^5 \\ &+ 398255520\theta^6 + 193439770\theta^7 - 405489800\theta^8 - 194447060\theta^9 + 229035400\theta^{10} \\ &+ 111485690\theta^{11} - 76620120\theta^{12} - 37823160\theta^{13} + 15162840\theta^{14} + 7551720\theta^{15} \\ &- 164520\theta^{16} - 82260\theta^{17} + 7560\theta^{18} + 3780\theta^{19} > 0,\end{aligned} \tag{A.34}$$

$$\mathcal{T} = (28 - 29\theta^2 + 5\theta^4)(42 - 88\theta^2 + 43\theta^4 - 6\theta^6) > 0. \tag{A.35}$$

Similarly, comparing the profit of firm 1 under data sharing to none,

$$\Pi_1^\star(1) - \Pi_1^\star(0) = \frac{\mathcal{A}^2\theta\mathcal{M}}{72\mathcal{T}^2}, \tag{A.36}$$

$$\begin{aligned}\mathcal{M} \equiv\ & 6435072 - 83381176\theta - 521216664\theta^2 + 55503336\theta^3 + 1308573312\theta^4 + 396597776\theta^5 \\ & - 1552180048\theta^6 - 800510000\theta^7 + 1022045552\theta^8 + 679055256\theta^9 - 399503388\theta^{10} \\ & - 318683236\theta^{11} + 93847728\theta^{12} + 88753584\theta^{13} - 12731880\theta^{14} - 14584920\theta^{15} \\ & + 880920\theta^{16} + 1303200\theta^{17} - 2160\theta^{18} - 4860\theta^{19} > 0.\end{aligned} \tag{A.37}$$

This establishes that both firms prefer firm 2 to share its data with firm 1. ■

Proof of Corollary 1. Comparing the equilibrium investment of firm 1 in market B under data sharing with its investment without data sharing, we find

$$\begin{aligned}v_1^\star(1) - v_1^\star(0) = & -\Big[2\mathcal{T}\Big]^{-1}\mathcal{A}(2+\theta)\theta\Big[140 - 176\theta + 352\theta^2 \\ & + 346\theta^3 - 475\theta^4 - 150\theta^5 + 168\theta^6 + 18\theta^7 - 18\theta^8\Big] < 0.\end{aligned} \tag{A.38}$$

Comparing these equilibrium investment choices for firm 2 gives

$$\begin{aligned}v_2^\star(1) - v_2^\star(0) = & -\Big[3\mathcal{T}\Big]^{-1}\mathcal{A}(2-\theta)\theta\Big[84 - 198\theta + 475\theta^2 \\ & + 1176\theta^3 - 13\theta^4 - 768\theta^5 - 186\theta^6 + 126\theta^7 + 42\theta^8)\Big] < 0.\end{aligned} \tag{A.39}$$

And lastly, comparing the equilibrium investment choices of firm 1 in market A gives

$$\begin{aligned}v_A^\star(1) - v_A^\star(0) = & \Big[3\mathcal{T}\Big]^{-1}\mathcal{A}\theta\Big[252 - 742\theta - 904\theta^2 \\ & + 566\theta^3 + 572\theta^4 - 133\theta^5 - 127\theta^6 + 9\theta^7 + 9\theta^8\Big].\end{aligned} \tag{A.40}$$

The numerator is positive if and only if $(252 - 742\theta - 904\theta^2 + 566\theta^3 + 572\theta^4 - 133\theta^5 - 127\theta^6 + 9\theta^7 + 9\theta^8) > 0$, which holds if $\theta < \theta^S = 0.270$ and does not otherwise. We now turn to output.

Comparing the equilibrium output of firm 1 in market B with and without data sharing gives

$$\begin{aligned} q_1^\star(1) - q_1^\star(0) = -\Big[6\mathcal{T}\Big]^{-1}\Big[&336 - 708\theta + 1720\theta^2 + 2857\theta^3 - 1714\theta^4 \\ &- 2137\theta^5 + 528\theta^6 + 582\theta^7 - 54\theta^8 - 54\theta^9\Big] < 0. \end{aligned} \tag{A.41}$$

Comparing these equilibrium output choices of firm 2 gives

$$\begin{aligned} q_2^\star(1) - q_2^\star(0) = \Big[\mathcal{T}\Big]^{-1}\mathcal{A}\theta\Big[&42 - 99\theta + 236\theta^2 + 396\theta^3 \\ &- 243\theta^4 - 305\theta^5 + 77\theta^6 + 85\theta^7 - 8\theta^8(1+\theta)\Big] > 0. \end{aligned} \tag{A.42}$$

And lastly, comparing the equilibrium output choices of firm 1 in market A gives

$$\begin{aligned} q_A^\star(1) - q_A^\star(0) = \Big[12\mathcal{T}\Big]^{-1}\mathcal{A}\theta\Big[&1008 - 24464\theta - 35444\theta^2 + 2128\theta^3 \\ &- 2563\theta^4 - 630\theta^5 - 696\theta^6 + 66\theta^7(1+\theta)\Big], \end{aligned} \tag{A.43}$$

which is positive if and only if $1008 - 24464\theta - 35444\theta^2 + 2128\theta^3 - 2563\theta^4 - 630\theta^5 - 696\theta^6 + 66\theta^7(1+\theta) > 0$, which in turn holds if $\theta < 0.307$ and does not otherwise. ■

Proof of Corollary 2. The result follows from the arguments made in the main text. ■

Proof of Proposition 2. The result follows from the arguments made in the main text. ■

References

Antón, M., Ederer, F., Giné, M. & Schmalz, M. (2023), 'Common ownership, competition, and top management incentives', *Journal of Political Economy* **131**(5), 1294–1355.

Argenton, C. & Prüfer, J. (2012), 'Search engine competition with network externalities', *Journal of Competition Law and Economics* **8**(1), 73–105.

Awaya, Y. & Krishna, V. (2016), 'On communication and collusion', *American Economic Review* **106**(2), 285–315.

Baldwin, C. Y. (2018), 'Design rules, volume 2: how technology shapes organizations', *Working Paper* .

Baysinger, B. & Hoskisson, R. E. (1989), 'Diversification strategy and R&D intensity in multiproduct firms', *Academy of Management Journal* **32**(2), 310–332.

Berger, P. G. & Ofek, E. (1995), 'Diversification's effect on firm value', *Journal of Financial Economics* **37**(1), 39–65.

Cason, T. N. & Mason, C. F. (1999), 'Information sharing and tacit collusion in laboratory duopoly markets', *Economic Inquiry* **37**(2), 258–281.

Cennamo, C. & Sokol, D. D. (2021), 'Can the EU regulate platforms without stifling innovation?', *Harvard Business Review Digital Articles* .

Choe, C., Cong, J. & Wang, C. (2022), 'Softening competition through unilateral sharing of customer data', *Management Science - forthcoming* .

Clarke, R. N. (1983), 'Collusion and the incentives for information sharing', *The Bell Journal of Economics* **14**(2), 383–394.

Crémer, J., de Montjoye, Y.-A. & Schweitzer, H. (2019), 'Competition policy for the digital era', *Report for the European Commission* .

Dubus, A. & Legros, P. (2022), 'The sale of data: Learning synergies before M&As', *Working Paper* .

Ederer, F. & Pellegrino, B. (2022), 'A tale of two networks: Common ownership and product market rivalry', *Working Paper* .

European Commission (2018), 'Study on data sharing between companies in Europe: Final report', https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en. Last accessed Nov 16, 2023.

European Commission (2020*a*), 'Business-to-Business data sharing: An economic and legal analysis'. Digital Economy Working Paper 2020-05.

European Commission (2020*b*), 'Digital Markets Act: Ensuring fair and open digital market', https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en. Last accessed Feb 23, 2022.

European Commission (2020*c*), 'A european strategy for data', https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066. Last accessed Nov 17, 2023.

European Commission (2022), 'Guidance on private sector data sharing', https://digital-strategy.ec.europa.eu/en/policies/private-sector-data-sharing. Last accessed Jan 19, 2023.

Fudenberg, D. & Tirole, J. (1984), 'The fat-cat effect, the puppy-dog ploy, and the lean and hungry look', *American Economic Review* **74**(2), 361–366.

Gal-Or, E. (1985), 'Information sharing in oligopoly', *Econometrica* **53**(2), 329–343.

Gilo, D., Moshe, Y. & Spiegel, Y. (2006), 'Partial cross ownership and tacit collusion', *The Rand Journal of Economics* **37**(1), 81–99.

Gomes, J. & Livdan, D. (2004), 'Optimal diversification: Reconciling theory and evidence', *The Journal of Finance* **59**(2), 507–535.

Graham, J. R., Lemmon, M. L. & Wolf, J. G. (2002), 'Does corporate diversification destroy value?', *The Journal of Finance* **57**(2), 695–720.

Gregory, R. W., Henfridsson, O., Kaganer, E. & Kyriakou, H. (2021), 'The role of artificial intelligence and data network effects for creating user value', *Academy of Management Review* **46**(3), 534–551.

Hagiu, A. & Wright, J. (2020), 'Data-enabled learning, network effects and competitive advantage', *The Rand Journal of Economics - forthcoming* .

Hill, C. W. & Hoskisson, R. E. (1987), 'Strategy and structure in the multiproduct firm', *Academy of Management Review* **12**(2), 331–341.

Huang, H., Parker, G., Tan, Y. R. & Xu, H. (2020), 'Altruism or shrewd business? Implications of technology openness on innovations and competition.', *MIS Quarterly* **44**(3).

Hunold, M. & Shekhar, S. (2022), 'Supply chain innovations and partial ownership', *Review of Industrial Organization* **60**(1), 109–145.

Jones, C. I. & Tonetti, C. (2020), 'Nonrivalry and the economics of data', *American Economic Review* **110**(9), 2819–58.

Kandori, M. & Matsushima, H. (1998), 'Private observation, communication and collusion', *Econometrica* **66**(3), 627–652.

Katz, M. L. & Shapiro, C. (1985), 'Network externalities, competition, and compatibility', *American Economic Review* **75**(3), 424–440.

Klein, T. J., Kurmangaliyeva, M., Prüfer, J., Prüfer, P. & Park, N. N. (2022), 'How important are user-generated data for search result quality? Experimental evidence', *Working Paper* .

Krämer, J. & Schnurr, D. (2022), 'Big data and digital markets contestability: Theory of harm and data access remedies', *Journal of Competition Law & Economics* **18**(2), 255–322.

Krämer, J. & Shekhar, S. (2022), 'Regulating algorithmic learning in digital platform ecosystems through data sharing and data siloing: Consequences for innovation and welfare', *Working Paper* .

Lang, L. H. & Stulz, R. M. (1994), 'Tobin's q, corporate diversification, and firm performance', *Journal of Political Economy* **102**(6), 1248–1280.

Lei, X., Chen, Y. & Sen, A. (2023), 'The value of external data for digital platforms: Evidence from a field experiment on search suggestions', *Working Paper* .

Lins, K. & Servaes, H. (1999), 'International evidence on the value of corporate diversification', *The Journal of Finance* **54**(6), 2215–2239.

O'Brien, D. P. & Salop, S. C. (1999), 'Competitive effects of partial ownership: Financial interest and corporate control', *Antitrust LJ* **67**, 559.

Parker, G., Petropoulos, G. & Van Alstyne, M. (2021), 'Platform mergers and antitrust', *Industrial and Corporate Change* **30**(5), 1307–1336.

Prüfer, J. & Schottmüller, C. (2021), 'Competing with big data', *The Journal of Industrial Economics* **69**(4), 967–1008.

Raith, M. (1996), 'A general model of information sharing in oligopoly', *Journal of Economic Theory* **71**(1), 260–288.

Schaefer, M., Sapi, G. & Lorincz, S. (2018), 'The effect of big data on recommendation quality: The example of internet search', *Working Paper* .

Shaked, A. & Sutton, J. (1990), 'Multiproduct firms and market structure', *The Rand Journal of Economics* **21**(1), 45–62.

Tucker, C. (2019), 'Digital data, platforms and the usual [antitrust] suspects: Network effects, switching costs, essential facility', *Review of Industrial Organization* **54**(4), 683–694.

Vives, X. (1984), 'Duopoly information equilibrium: Cournot and Bertrand', *Journal of Economic Theory* **34**(1), 71–94.

Appendix B. Proofs for the extensions

In this appendix we provide proofs and supporting calculations of the extensions in Section 5.

B.1. Differentiated price competition

The model description is available in Section 5.1. We proceed by presenting the analysis first in the case that the specialist firm 2 does not share its data, and then when it does.

Specialist firm does not share data

At stage 3, given expectations, consumers act and their demands are

$$q_1(v_1, v_2, p_1, p_2, q_A^e) = \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2) + \theta q_A^e}{2t}, \tag{B.1}$$

$$q_2(v_2, v_1, p_2, p_1, q_A^e) = 1 - q_1(\cdot), \tag{B.2}$$

$$q_A(v_A, p_A, q_1^e) = v_A + \theta q_1^e - p_A. \tag{B.3}$$

In any equilibrium, consumers' expectations should match the outcome and so we set $q_A^e = q_A$, $q_1^e = q_1$ and solve for demands as functions of price and investment levels.

$$q_1(v_1, v_2, v_A, p_1, p_2, p_A) = \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2) + \theta(v_A - p_A)}{2t - \theta^2}, \tag{B.4}$$

$$q_2(v_2, v_1, v_A, p_2, p_1, p_A) = 1 - q_1(\cdot), \tag{B.5}$$

$$q_A(v_A, v_1, v_2, p_A, p_1, p_2) = \frac{2t(v_A - p_A) + \theta(t + v_1 - v_2 - (p_1 - p_2))}{2t - \theta^2}. \tag{B.6}$$

Substituting these demands in the profit expression yields

$$\Pi_1 = \underbrace{p_A q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1 q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2 q_2(\cdot) - I(v_2). \tag{B.7}$$

Differentiating the profit of the generalist firm with respect to its prices p_1 and p_A and the specialist firm with respect to p_2 gives a system of first order conditions. Solving those yields prices as functions of investment levels given as follows.

$$p_1(v_1, v_2, v_A) = t + \frac{2(v_1 - v_2) - \theta(v_A + 2\theta)}{6}, \tag{B.8}$$

$$p_2(v_2, v_1, v_A) = t + \frac{v_2 - v_1 - \theta(v_A + 2\theta)}{6}, \tag{B.9}$$

$$p_A(v_A, v_1, v_2) = \frac{v_A}{2}. \tag{B.10}$$

Substituting these prices into the profit expressions yields profits as functions of investments.

$$\Pi_1 = \underbrace{p_A(\cdot)q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1(\cdot)q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2(\cdot)q_2(\cdot) - I(v_2). \tag{B.11}$$

Differentiating Π_1 with respect to its prices, v_1 and v_A and Π_2 with respect to v_2, and solving the resulting system of first order conditions yields

$$v_1^\star(0) = \frac{4 - 18t + 6\theta^2}{12 - 54t + 39\theta^2}, \; v_2^\star(0) = \frac{4 - 18t + 20\theta^2}{12 - 54t + 39\theta^2}, \; v_A^\star(0) = \frac{4\theta(2 - 9t + 3\theta^2)}{12 - 54t + 39\theta^2}. \tag{B.12}$$

Substituting these investment levels into prices, demands and profits we find

$$p_1^\star(0) = v_1^\star(0)\left(3t - \frac{5\theta^2}{2}\right), \; p_2^\star(0) = 3v_2^\star(0)\left(t - \frac{\theta^2}{2}\right), \; p_A^\star(0) = \frac{v_A^\star(0)}{2}, \tag{B.13}$$

$$q_1^\star(0) = \frac{3v_1^\star(0)}{2}, \; q_2^\star(0) = \frac{3v_2^\star(0)}{2}, \; q_A^\star(0) = \frac{5v_A^\star(0)}{4}, \tag{B.14}$$

$$\Pi_1^\star(0) = \frac{v_A^\star(0)^2}{8} + \frac{v_1^\star(0)^2}{2}\left(9t - \frac{15\theta^2}{2} - 1\right), \; \Pi_2^\star(0) = \frac{v_2^\star(0)^2}{2}\left(9t - \frac{9\theta^2}{2} - 1\right). \tag{B.15}$$

Specialist firm shares data

At stage 3, given expectations, consumers act and their demands are

$$q_1(v_1, v_2, p_1, p_2, q_A^e) = \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2) + \theta q_A^e}{2t}, \tag{B.16}$$

$$q_2(v_2, v_1, p_2, p_1, q_A^e) = 1 - q_1(\cdot), \tag{B.17}$$

$$q_A(v_A, p_A, q_1^e, q_2^e) = v_A + \theta(q_1^e + q_2^e) - p_A. \tag{B.18}$$

In any equilibrium, consumers' expectations should match the outcome and so we set $q_A^e = q_A$, $q_1^e = q_1$ and $q_2^e = q_2$ and solve for demands as functions of price and investment levels.

$$q_1(v_1, v_2, v_A, p_1, p_2, p_A) = \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2) + \theta(\theta + v_A - p_A)}{2t - \theta^2}, \tag{B.19}$$

$$q_2(v_2, v_1, v_A, p_2, p_1, p_A) = 1 - q_1(\cdot), \tag{B.20}$$

$$q_A(v_A, p_A) = \theta + v_A - p_A. \tag{B.21}$$

Substituting these demands in the profit expressions yields

$$\Pi_1 = \underbrace{p_A q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1 q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2 q_2(\cdot) - I(v_2). \tag{B.22}$$

Differentiating the profit of the generalist firm with respect to its prices p_1 and p_A and the specialist firm with respect to p_2 gives a system of first order conditions. Solving those yields prices as functions of investment levels given as follows.

$$p_1(v_1, v_2, v_A) = \frac{2t(6t + 2(v_1 - v_2) + \theta(v_A + \theta))}{12t - \theta^2}, \tag{B.23}$$

$$p_2(v_2, v_1, v_A) = \frac{2t(6t - 2(v_1 - v_2) - \theta(v_A + 2\theta))}{12t - \theta^2}, \tag{B.24}$$

$$p_A(v_A, v_1, v_2) = v_A + \theta - \frac{6tv_A + \theta(9t + v_1 - v_2)}{12t - \theta^2}. \tag{B.25}$$

Substituting these prices into the profit expressions yields profits as functions of investments.

$$\Pi_1 = \underbrace{p_A(\cdot) q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1 q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2 q_2(\cdot) - I(v_2). \tag{B.26}$$

Differentiating Π_1 with respect to its prices, v_1 and v_A and Π_2 with respect to v_2, and solving the resulting system of first order conditions yields

$$v_1^\star(1) = \frac{(8t-\theta^2)(4t(9t-2)+\theta^2(9t-\theta^2))}{(12t-\theta^2)(72t^2+\theta^2+\theta^4-16t(1+\theta^2))}, \tag{B.27}$$

$$v_2^\star(1) = \frac{8t(4t(9t-2)-\theta^2(25t-1-2\theta^2))}{(12t-\theta^2)(72t^2+\theta^2+\theta^4-16t(1+\theta^2))}, \tag{B.28}$$

$$v_A^\star(1) = \frac{128t^2\theta(9t-2)+\theta^5(11t-1)-12t\theta^3(19t-3)}{(12t-\theta^2)(72t^2+\theta^2+\theta^4-16t(1+\theta^2))}. \tag{B.29}$$

Substituting these investment levels into prices, demands and profits we find

$$p_1^\star(1) = v_1^\star(1)2t\left(1+\frac{4t}{8t-\theta^2}\right),\ p_2^\star(1) = v_2^\star(1)\left(3t-\frac{\theta^2}{4}\right), \tag{B.30}$$

$$p_A^\star(1) = \frac{88t^2\theta(9t-2)+\theta^5(30t-1)-\theta^7-2t\theta^3(141t-14)}{(12t-\theta^2)(72t^2+\theta^2+\theta^4-16t(1+\theta^2))}, \tag{B.31}$$

$$q_1^\star(1) = v_1^\star(1)\left(1+\frac{4t}{8t-\theta^2}\right),\ q_2^\star(1) = v_2^\star(1)\left(\frac{3}{2}-\frac{\theta^2}{8t}\right), \tag{B.32}$$

$$q_A^\star(1) = \frac{136t^2\theta(9t-2)+\theta^5(9t-1)-6t\theta^3(35t-6)}{(12t-\theta^2)(72t^2+\theta^2+\theta^4-16t(1+\theta^2))}, \tag{B.33}$$

$$\Pi_1^\star(1) = p_A^\star(1)q_A^\star(1) - \frac{v_A^\star(1)^2}{2} + \frac{v_1^\star(1)^2}{2}\left(\frac{4t(12t-\theta^2)^2}{(8t-\theta^2)^2}-1\right), \tag{B.34}$$

$$\Pi_2^\star(1) = \frac{(v_2^\star(1))^2}{2}\left(9t-\frac{3\theta^2}{2}+\frac{\theta^4}{16t}-1\right). \tag{B.35}$$

Comparing firm 2's profit under data sharing versus when firm 2 does not share data, we find

$$\Pi_2^\star(1) - \Pi_2^\star(0) > 0. \tag{B.36}$$

Firm 2 finds it profitable to share data with firm 1 under our regularity Assumption 1.

B.2. Intra-market externalities

We consider the following demand system, which includes an intra-market externality.

$$P_A(v_A, q_1, \Phi q_2, q_A) = \mathcal{A} + v_A + \theta(q_1 + \Phi q_2) - \beta_A q_A, \tag{B.37}$$

$$P_1(v_1, q_A, q_1, Q_B) = \mathcal{B} + v_1 + \theta q_A + \sigma q_1 - \beta_B Q_B, \tag{B.38}$$

$$P_2(v_2, q_2, Q_B) = \mathcal{B} + v_2 + \sigma q_2 - \beta_B Q_B. \tag{B.39}$$

We first provide solutions for quantity and investment choices with any values $\theta, \sigma \geq 0$ so that they provide well-defined solutions, which, as in our main model, means we assume they are sufficiently small. Otherwise, the model is unchanged and profits are

$$\Pi_1 = P_A(\cdot)q_A + P_1(\cdot)q_1 - I(v_A) - I(v_1), \tag{B.40}$$

$$\Pi_2 = P_2(\cdot)q_2 - I(v_2). \tag{B.41}$$

At stage 3, firm 1 chooses q_A, q_1 to maximize Π_1 and firm 2 chooses q_2 to maximize Π_2. Solving the resulting system of equations yields the optimal output choices as functions of θ and σ.

$$\begin{aligned}\hat{q}_A(\cdot) = [2\rho]^{-1}\Big[&\alpha((\Phi+2)\theta+6) - \sigma(\Phi\theta(\alpha+v_2+1) + 8(\alpha+v_A) + 2\theta(\alpha+v_1+1)) \\ &+ \theta(\Phi(-v_1) + 2(\Phi-1)v_2 + \Phi + 4v_1 + 2) - 8\sigma + 2\sigma^2(\alpha+v_A+1) + 6v_A + 6\Big],\end{aligned} \tag{B.42}$$

$$\hat{q}_1(\cdot) = [2\rho]^{-1}\Big[(\Phi\theta^2 - 4)(\alpha+v_2+1) - 2(\sigma-2)(\theta(\alpha+v_A+1) + 2(\alpha+v_1+1))\Big], \tag{B.43}$$

$$\hat{q}_2(\cdot) = \rho^{-1}\Big[2 - \alpha(\theta^2+\theta+2\sigma-2) - 2\sigma - \theta(\theta+v_A+1) - 2v_1 - v_2(\theta^2+2\sigma-4)\Big], \tag{B.44}$$

$$\rho = \theta^2(\Phi + 2\sigma - 4) + 4(\sigma-3)(\sigma-1). \tag{B.45}$$

We next solve the game at stage 2. Firms choose investment to maximize profits, which yields

$$\begin{aligned}v_A^*(\Phi) = \Big[2\psi\Big]^{-1}\Big[&(\alpha+1)(\theta^5(\Phi^2\sigma(5-2\sigma)-2\Phi(\sigma-2)^2(2\sigma+1)-4(\sigma-2)^2(2\sigma-1))\\&+2\theta^3(\Phi^2(2-(\sigma-2)\sigma(2\sigma-5))+\Phi(2\sigma(21-4\sigma((\sigma-6)\sigma+11))+8)\\&-8(\sigma-2)(2\sigma-3)((\sigma-3)\sigma+1))+2\theta^4(\sigma-2)(\Phi(\sigma(2\sigma-7)+2)+\sigma(5-2\sigma)^2\\&-4)+4\theta^2(\Phi(\sigma(\sigma(\sigma(2\sigma-15)+38)-36)+10)+2(\sigma-2)(\sigma-1)(\sigma(4(\sigma-6)\sigma\\&+39)-8))-8\theta((\sigma-2)\sigma(2\sigma-5)-2)(\Phi(\sigma-3)(\sigma-1)+2(\sigma-2)^2)\\&+8((\sigma-2)\sigma(2\sigma-5)-2)(\sigma(\sigma(2\sigma-13)+26)-14))\Big],\end{aligned} \tag{B.46}$$

$$\begin{aligned}v_1^*(\Phi) = \Big[2\psi\Big]^{-1}\Big[&(\alpha+1)(2\theta^4(\Phi^3(\sigma-1)+2\Phi^2(\sigma(3\sigma-11)+9)+\Phi(\sigma-2)(\sigma(8\sigma-35)\\&+31)-3(\sigma-2)^2(2\sigma-1))+4\theta^2 2\Phi^2(\sigma-3)(\sigma-1)^2+\Phi(\sigma-1)(4\sigma-15)(\sigma\\&-2)^2-6(2\sigma-3)((\sigma-3)\sigma+1)(\sigma-2)+\Phi\theta^6(\Phi+2\sigma-4)^2\\&-2\theta^5(\sigma-2)(\Phi+2\sigma-4)(\Phi+2\sigma-3)-4\theta^3(\sigma-1)(\Phi+2\sigma-4)(\Phi(\sigma-3)\\&+2(\sigma-2)(2\sigma-5))-8\theta(\Phi(\sigma(\sigma(\sigma(2\sigma-15)+38)-36)+10)+2(\sigma(2\sigma-7)\\&+4)(\sigma-2)^3)-24(\sigma-2)^2((\sigma-2)\sigma(2\sigma-5)-2))\Big],\end{aligned} \tag{B.47}$$

$$\begin{aligned}v_2^*(\Phi) = \psi^{-1}\Big[&-((\alpha+1)(\sigma-2)(\theta^2+2\sigma-4)(2\theta^4(\Phi+2\sigma-4)+2\theta^3(\Phi+2\sigma-4)\\&+\theta^2(4\Phi(\sigma-1)+\sigma(14\sigma-47)+32)-2\theta(\Phi-4(\sigma-3)(\sigma-1))\\&+6((\sigma-2)\sigma(2\sigma-5)-2))\Big],\end{aligned} \tag{B.48}$$

$$\begin{aligned}\psi \equiv\ &(\theta^6(\Phi+2\sigma-4)(\Phi^2+4\Phi(\sigma-2)+2(\sigma-2)(2\sigma-3))+\theta^4(2\Phi^2(\sigma(6\sigma-25)\\&+20)-\Phi^3+\Phi(\sigma-2)(\sigma(46\sigma-173)+118)+(\sigma-2)(\sigma(4\sigma(11\sigma-59)+379)\\&-172))+2\theta^2(-2\Phi^2(\sigma-3)(\sigma-1)+\Phi(\sigma(\sigma(\sigma(22\sigma-169)+442)-444)\\&+142)+2(\sigma-2)(2\sigma-1)(\sigma(\sigma(10\sigma-67)+141)-88))\\&+12((\sigma-2)\sigma(2\sigma-5)-2)(\sigma(\sigma(2\sigma-13)+26)-14)).\end{aligned} \tag{B.49}$$

Substituting investment choices (B.46)-(B.48) into (B.42)-(B.44) provides the optimal output choices, $q_m^*(\Phi)$ for $m \in \{A, 1, 2\}$, as in the main analysis. Calculating, we replicate Corollary 1:

$$v_1^\star(1) < v_1^\star(0),\ q_1^\star(1) < q_1^\star(0) \quad \text{and} \quad v_2^\star(1) > v_2^\star(0),\ q_2^\star(1) > q_2^\star(0). \tag{B.50}$$

We also confirm that when firm 2 shares its data, firm 1 accommodates it less when σ is higher:

$$\frac{\partial v_1^\star(1)}{\partial \sigma} < 0, \quad \frac{\partial q_1^\star(1)}{\partial \sigma} < 0. \tag{B.51}$$

We substitute investment choices (B.46)-(B.48) into (B.42)-(B.44) and first consider $\theta = \sigma$:

$$\begin{aligned}\Pi_2^*(\Phi) = -\Big[&4(\theta(\theta(\Phi^3\theta^2(\theta^2-1)+2\Phi^2(3\theta^2-1)(\theta^3-8\theta+6)+\Phi(3\theta(\theta(\theta(4\theta^3-59\theta\\&+42)+216)-296)+284)+\theta(\theta(\theta(\theta(8\theta^3-164\theta+115)+1090)-1448)\\&-1864)+4240)-2304)+336)^2\Big]^{-1}\Big[(\alpha+1)^2(\theta-2)(\theta(\theta^3(\Phi^2-60)+\theta^4(4\Phi\\&+2)+8\theta^2(3-4\Phi)+24\theta(\Phi+10)+4\theta^5-288)+80)(\theta(2\Phi(\theta(\theta(\theta+3)-2)\\&-1)+\theta(\theta(2\theta(2\theta+5)-35)-54)+84)-12)^2\Big].\end{aligned} \tag{B.52}$$

It is then straightforward to confirm that $\Pi_2^*(1) > \Pi_2^*(0)$. Second, we consider when $\sigma > \theta$:

$$\begin{aligned}\Pi_2^*(\Phi) = -\Big[&4(\theta^6(\Phi+2\sigma-4)(\Phi^2+4\Phi(\sigma-2)+2(\sigma-2)(2\sigma-3))\\&+\theta^4(2\Phi^2(\sigma(6\sigma-25)+20)-\Phi^3+\Phi(\sigma-2)(\sigma(46\sigma-173)+118)\\&+(\sigma-2)(\sigma(4\sigma(11\sigma-59)+379)-172))+2\theta^2(-2\Phi^2(\sigma-3)(\sigma-1)\\&+\Phi(\sigma(\sigma(\sigma(22\sigma-169)+442)-444)+142)+2(\sigma-2)(2\sigma-1)(\sigma(\sigma(10\sigma\\&-67)+141)-88))+12((\sigma-2)\sigma(2\sigma-5)-2)(\sigma(\sigma(2\sigma-13)+26)\\&-14))^2\Big]^{-1}\Big[(\alpha+1)^2(\sigma-2)(2\theta^4(\Phi+2\sigma-4)+2\theta^3(\Phi+2\sigma-4)+\theta^2(4\Phi\\&(\sigma-1)+\sigma(14\sigma-47)+32)-2\theta(\Phi-4(\sigma-3)(\sigma-1))+6((\sigma-2)\sigma(2\sigma-5)\\&-2))^2(\theta^4(\Phi^2+4\Phi(\sigma-2)+2(\sigma-2)(2\sigma-3))+8\theta^2(\Phi(\sigma-3)(\sigma-1)\\&+(\sigma-2)(\sigma(2\sigma-7)+4))+8\sigma(\sigma(\sigma(2\sigma-15)+38)-36)+80)\Big].\end{aligned} \tag{B.53}$$

And here too $\Pi_2^*(1) > \Pi_2^*(0)$, i.e., firm 2 prefers to share its data ($\Phi = 1$) than not ($\Phi = 0$). In addition, one can confirm that firm 2's profits are falling in σ because $\frac{d\Pi_2^*(\Phi)}{d\sigma} < 0$ for $\Phi \in \{0, 1\}$.

B.3. Heterogeneous cross-market externalities

Throughout, θ_A and θ_B satisfy the same condition as in our main model: $\theta_A, \theta_B < \overline{\theta} \approx 0.353$. We consider the following demand system

$$P_A(v_A, q_1, \Phi q_2, q_A) = \mathcal{A} + v_A + \theta_B(q_1 + \Phi q_2) - \beta_A q_A, \tag{B.54}$$

$$P_1(v_1, q_A, q_1, Q_B) = \mathcal{B} + v_1 + \theta_A q_A - \beta_B Q_B, \tag{B.55}$$

$$P_2(v_2, q_2, Q_B) = \mathcal{B} + v_2 - \beta_B Q_B. \tag{B.56}$$

Profits are defined in the usual way, but with demands (B.54)-(B.56). Optimal outputs are

$$\begin{aligned}\hat{q}_A(\cdot) = \rho^{-1}\Big[&6 + 6v_A + (1 - v_1(-2+\Phi) + \Phi + v_2(-1+2\Phi))\theta_A + \theta_B + 2v_1\theta_B - v_2\theta_B \\ &+ \alpha(6 + \theta_A + \Phi\theta_A + \theta_B)\Big],\end{aligned} \tag{B.57}$$

$$\begin{aligned}\hat{q}_1(\cdot) = \Big[2\rho\Big]^{-1}\Big[&16(1+v_1+\alpha) + 4(1+v_A+\alpha)(\theta_A+\theta_B) + (1+v_2+\alpha)(-8 \\ &+ \Phi\theta_A(\theta_A+\theta_B)],\end{aligned} \tag{B.58}$$

$$\begin{aligned}\hat{q}_2(\cdot) = -\Big[2\rho\Big]^{-1}\Big[&8v_1 + v_2(-4+\theta_A+\theta_B)(4+\theta_A+\theta_B) + (\theta_A+\theta_B)^2 \\ &+ \alpha(-2+\theta_A+\theta_B)(4+\theta_A+\theta_B) + 2(-4+\theta_A+v_A\theta_A+\theta_B+v_A\theta_B)\Big],\end{aligned} \tag{B.59}$$

$$\rho \equiv 24 + ((-2+\Phi)\theta_A - 2\theta_B)(\theta_A+\theta_B). \tag{B.60}$$

Using these equilibrium outputs, we solve for the optimal investments of each firm at stage 2:

$$\begin{aligned}v_A^*(\Phi) = \psi^{-1}\Big[&2(1+\alpha)(448 - (-1+\Phi)\theta_A^5 + \theta_A^4(2+5\theta_B - 2\Phi(1+2\theta_B)) \\ &+ 2\theta_A^2(20\Phi + 2\Phi(4+\Phi)\theta_B - 3(-2+\Phi)\theta_B^2 + (5-2\Phi)\theta_B^3 - 4(8+9\theta_B)) \\ &+ \theta_B(128 + \theta_B(4+\theta_B)(-16 + (-2+\theta_B)\theta_B)) + 2\theta_A^3(2\Phi^2 + (2+\theta_B)(-6+5\theta_B) \\ &+ \phi(4 - 3\theta_B(1+\theta_B))) + \theta_A(128 - \Phi(-4+\theta_B)(24+\theta_B(16+\theta_B(6+\theta_B))) \\ &+ \theta_B(-128 + \theta_B(-72 + \theta_B(8+5\theta_B))))\Big]\end{aligned} \tag{B.61}$$

$$
\begin{aligned}
v_1^*(\Phi) = \Big[2\psi\Big]^{-1}\Big[&(1+\alpha)((-2+\Phi)^2\Phi\theta_A^6 + \theta_A^4(12-4\Phi(31+2(-9+\Phi)\Phi) \\
&+4(-3+\Phi)(-5+3\Phi)\theta_B + \Phi(40+3(-8+\Phi)\Phi)\theta_B^2) + \theta_A^3(-320+16(16 \\
&-3\Phi)\Phi + 48\theta_B - 4\Phi(93+2(-18+\Phi)\Phi)\theta_B + 12(-5+\Phi)(-2+\Phi)\theta_B^2 + \Phi(40 \\
&+(-16+\Phi)\Phi)\theta_B^3) + (-2+\Phi)\theta_A^5(-6+\Phi(4+(-10+3\Phi)\theta_B)) + 4(-4+\theta_B) \\
&(4+\theta_B)(-24+\theta_B(-32+3\theta_B(1+\theta_B))) + 4\theta_A^2(6(3+\theta_B)(-4+\theta_B(-12+5\theta_B)) \\
&-\Phi^2(-4+\theta_B)(-12+\theta_B(-6+\theta_B(3+\theta_B))) + \Phi(240+\theta_B(128+\theta_B(-93 \\
&+\theta_B(-14+5\theta_B)))) + 2\theta_A(1024+6\theta_B(-48+\theta_B(-80+\theta_B(4+5\theta_B))) \\
&+\Phi(-320+\theta_B(480+\theta_B(128+\theta_B(-62+\theta_B(-7+2\theta_B)))))))\Big]
\end{aligned}
\tag{B.62}
$$

$$
\begin{aligned}
v_2^*(\Phi) = -\psi^{-1}\Big[&(1+\alpha)(-4+\theta_A+\theta_B)(4+\theta_A+\theta_B)(\theta_A(48-2\theta_A(-16+\theta_A(2+\theta_A)) \\
&+\Phi(-8+(-2+\theta_A)\theta_A(4+\theta_A)) + 48(-1+\theta_B) + \theta_A(64-4\theta_A(3+2\theta_A) \\
&+\Phi(-8+\theta_A(4+3\theta_A))\theta_B + (32+2(-6+\Phi)\theta_A + 3(-4+\Phi)\theta_A^2)\theta_B^2 \\
&+(-4+(-8+\Phi)\theta_A)\theta_B^3 - 2\theta_B^4)\Big]
\end{aligned}
\tag{B.63}
$$

$$
\begin{aligned}
\psi \equiv\ & 2688 + (-3+\Phi)(-2+\Phi)(-1+\Phi)\theta_A^6 + (-36+\Phi(55+3(-8+\Phi)\Phi))\theta_A^5\theta_B \\
&-1408\theta_B^2 + 172\theta_B^4 - 6\theta_B^6 + \theta_A^3\theta_B(688-4\Phi(177+(-40+\Phi)\Phi) \\
&+(-4+\Phi)(30+(-20+\Phi)\Phi)\theta_B^2) + \theta_A^4(172-4\Phi(59+(-20+\Phi)\Phi) \\
&+(-90+\Phi(110+3(-12+\Phi)\Phi))\theta_B^2) + \theta_A\theta_B(16(-176+71\Phi) \\
&+4(172-59\Phi)\theta_B^2 + (-36+11\Phi)\theta_B^4) + \theta_A^2(-16(88+\Phi(-71+6\Phi)) \\
&+4(258+\Phi(-177+20\Phi))\theta_B^2 + (-90+(55-6\Phi)\Phi)\theta_B^4).
\end{aligned}
\tag{B.64}
$$

Substituting investment choices (B.61)-(B.64) into (B.57)-(B.59) provides the optimal output choices, $q_m^*(\Phi)$ for $m \in \{A, 1, 2\}$, as in the main analysis. Calculating, we replicate Corollary 1:

$$
v_1^\star(1) < v_1^\star(0),\ q_1^\star(1) < q_1^\star(0) \quad \text{and} \quad v_2^\star(1) > v_2^\star(0),\ q_2^\star(1) > q_2^\star(0). \tag{B.65}
$$

In turn, the profits of firm 2 when it shares data ($\Phi = 1$) and when it does not ($\Phi = 0$) are

$$\begin{aligned}\Pi_2^*(1) = -\Big[&4((41\theta_A^2 - 172)\theta_B^4 + \theta_A(33\theta_A^2 - 452)\theta_B^3 + (\theta_A - 2)(\theta_A + 2)(13\theta_A^2 - 352)\theta_B^2 \\ &- 4(\theta_A^2 - 12)(3\theta_A^2 - 56) + 2\theta_A(\theta_A^4 - 68\theta_A^2 + 840)\theta_B + 25\theta_A\theta_B^5 + 6\theta_B^6)^2\Big]^{-1} \\ &\Big[(1+\alpha)^2(\theta_A^4 + \theta_A^3(5\theta_B + 2) + \theta_A^2(\theta_B(9\theta_B + 8) - 24) + \theta_A(\theta_B(\theta_B(7\theta_B + 10) - 56) \\ &- 40) + 2\theta_B(\theta_B(\theta_B(\theta_B + 2) - 16) - 24) + 48)(2((\theta_A + \theta_B)(\theta_A + 2\theta_B) - 24) \\ &(\theta_A^5\theta_B + \theta_A^4(\theta_B(7\theta_B - 2) - 4) + \theta_A^3(\theta_B(\theta_B(19\theta_B - 10) - 72) + 16) \\ &+ \theta_A^2(\theta_B(\theta_B(\theta_B(25\theta_B - 18) - 252) + 104) + 176) + 2\theta_A(\theta_B(\theta_B(\theta_B(\theta_B(8\theta_B - 7) \\ &- 154) + 84) + 592) - 160) + 4(\theta_B^2 - 12)(\theta_B(\theta_B((\theta_B - 1)\theta_B - 19) + 8) + 48)) \\ &+ 8\theta_A^2(\theta_A(\theta_A(\theta_A(5\theta_A + 4) - 244) - 144) + 3744) - 2(81\theta_A + 4)\theta_B^7 \\ &- 2(\theta_A(199\theta_A + 22) - 568)\theta_B^6 + 2(\theta_A(2404 - \theta_A(269\theta_A + 50)) + 144)\theta_B^5 \\ &- 8(\theta_A(\theta_A(3\theta_A(18\theta_A + 5) - 1016) - 138) + 2004)\theta_B^4 - 2(\theta_A(\theta_A(\theta_A(\theta_A(103\theta_A \\ &+ 40) - 3464) - 808) + 21840) + 1664)\theta_B^3 - 2(\theta_A(\theta_A(\theta_A(\theta_A(\theta_A(27\theta_A + 14) \\ &- 1532) - 552) + 20800) + 3776) - 44416)\theta_B^2 - 2\theta_A(\theta_A(\theta_A + 8)(\theta_A(\theta_A(\theta_A(3\theta_A \\ &- 22) - 140) + 952) + 336) - 57088)\theta_B + 2048(5\theta_A + 6\theta_B - 69) - 28\theta_B^8)\Big].\end{aligned} \tag{B.66}$$

$$\begin{aligned}\Pi_2^*(0) = \Big[&2((3(\theta_A + \theta_B)^2 - 86)(\theta_A + \theta_B)^4 + 704(\theta_A + \theta_B)^2 - 1344)^2\Big]^{-1} \\ &\Big[(1+\alpha)^2((\theta_A + \theta_B)^2 - 8)(3(\theta_A + \theta_B)^2 - 40)(\theta_A^4 + \theta_A^3(4\theta_B + 2) \\ &+ 2\theta_A^2(3\theta_B(\theta_B + 1) - 8) + \theta_A(2\theta_B(\theta_B(2\theta_B + 3) - 16) - 24) \\ &+ \theta_B(\theta_B(\theta_B(\theta_B + 2) - 16) - 24) + 24)^2\Big],\end{aligned} \tag{B.67}$$

Comparing confirms $\Pi_2^*(1) > \Pi_2^*(0)$, i.e., that data sharing is profitable for firm 2.

B.4. Two generalist firms

We consider the following demand system and profits:

$$P_{A1}(v_{A1}, q_{B1}, \Phi_1 q_2, Q_A) = \mathcal{A} + v_{A1} + \theta(q_{B1} + \Phi_1 q_2) - \beta_A Q_A, \quad \text{(B.68)}$$

$$P_{B1}(v_{B1}, q_{A1}, Q_B) = \mathcal{B} + v_{B1} + \theta q_{A1} - \beta_B Q_B, \quad \text{(B.69)}$$

$$P_2(v_2, q_2, Q_B) = \mathcal{B} + v_2 - \beta_B Q_B, \quad \text{(B.70)}$$

$$P_{A3}(v_{A3}, q_{B3}, \Phi_3 q_2, Q_A) = \mathcal{A} + v_{A3} + \theta(q_{B3} + \Phi_3 q_2) - \beta_A Q_A, \quad \text{(B.71)}$$

$$P_{B3}(v_{B3}, q_{A3}, Q_B) = \mathcal{B} + v_{B3} + \theta q_{A3} - \beta_B Q_B. \quad \text{(B.72)}$$

$$\Pi_1 = P_{A1}(\cdot) q_{A1} + P_{B1}(\cdot) q_{B1} - I(v_{A1}) - I(v_{B1}), \quad \text{(B.73)}$$

$$\Pi_2 = P_2(\cdot) q_2 - I(v_2), \quad \text{(B.74)}$$

$$\Pi_3 = P_{A3}(\cdot) q_{A3} + P_{B3}(\cdot) q_{B3} - I(v_{A3}) - I(v_{B3}). \quad \text{(B.75)}$$

Solving for optimal outputs and investments, we find that each generalist firm chooses less when firm 2 shares its data, as in Corollary 1. For profits, we find the following relationships:

$$\Pi_2^*(1,1) > \Pi_2^*(0,1) = \Pi_2^*(1,0) > \Pi_2^*(0,0). \quad \text{(B.76)}$$

$$\Pi_1^*(1,0) > \Pi_1^*(0,0) \text{ and } \Pi_3^*(0,1) > \Pi_3^*(0,0). \quad \text{(B.77)}$$

$$\Pi_1^*(1,1) > \Pi_1^*(0,1) \text{ and } \Pi_3^*(1,1) > \Pi_3^*(1,0). \quad \text{(B.78)}$$

These establish that firm 2 shares with both 1 and 3. Profits per sharing arrangement are:

$$\begin{aligned}\Pi_1^*(0,0) = \Pi_3^*(0,0) = -[8\kappa]^{-1}\Big[&(\alpha+1)^2(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(4\theta(2\theta(6\theta+11)\\ &-301)-2077)+12422)+20061)-67212)-101641)+200430)\\ &+286721)-312620)-433050)+207660)+288174)-22644)-42768)\Big]\end{aligned} \quad \text{(B.79)}$$

$$\begin{aligned}\Pi_2^*(0,0) = [2\kappa]^{-1}\Big[&3(\alpha+1)^2(\theta^2-7)(\theta^2-3)(\theta(\theta(\theta(\theta(2\theta(\theta+2)-17)-28)+34)\\ &+29)-7)^2\Big];\end{aligned} \quad \text{(B.80)}$$

$$\kappa \equiv (6\theta^8 - 91\theta^6 + 458\theta^4 - 800\theta^2 + 273)^2 \quad \text{(B.81)}$$

$$
\begin{aligned}
\Pi_1^*(1,0) = \Pi_3^*(0,1) = -\Big[2\lambda\Big]^{-1}\Big[&(\alpha+1)^2(\theta \\
&(\theta(\theta(3\theta(\theta(24\theta(\theta(90\theta(3\theta+14)-13429)-67267)+6962381)+38391720) \\
&-252775544)-1602965656)+1813130178)+14408466752) \\
&-7046695432)-86974486664)+5237400952)+354952456272) \\
&+85791045984)-962052218848)-439656822272)+1659630641856) \\
&+1007601439296)-1690457197056)-1197788585984)+881795848192) \\
&+682442820864)-171469357056)-138027108352)+11373232128) \\
&+8766640128)-226492416)-175177728)\Big],
\end{aligned}
\tag{B.82}
$$

$$
\begin{aligned}
\Pi_2^*(1,0) = \Pi_2^*(0,1) = \lambda^{-1}\Big[&2(\alpha+1)^2(\theta^2-6)(5\theta^2-42)(\theta(\theta(\theta(\theta(3\theta(\theta(\theta(\theta(2\theta(\theta(12\theta(\theta \\
&+2)-233)-431)+3119)+4981)-8404)-9992)+25544)+19596) \\
&-5688)-912)+224)^2\Big],
\end{aligned}
\tag{B.83}
$$

$$
\begin{aligned}
\Pi_1^*(0,1) = \Pi_3^*(1,0) = -\Big[2\lambda\Big]^{-1}\Big[&(\alpha+1)^2(\theta(3\theta(\theta \\
&(12\theta(4\theta(3\theta(15\theta(8\theta+19)-6239)-44066)+1715869)+47997391) \\
&-273064898)-1888276739)+6936053756)+15789630576) \\
&-39015358288)-87635893696)+147016689824)+325911606496) \\
&-365124375392)-801520526864)+571895310912)+1257732024448) \\
&-516311126784)-1177878527488)+218371945984)+578839090432) \\
&-19558063104)-113650075648)+722264064)+7808765952) \\
&+1622016)-175177728)\Big],
\end{aligned}
\tag{B.84}
$$

$$
\begin{aligned}
\lambda \equiv \Big(&-360\theta^{14}+106602\theta^{12}-1225113\theta^{10}+6916666\theta^{8}-19173552\theta^{6}+22632264\theta^{4} \\
&-8095568\theta^{2}+34944\Big)^2
\end{aligned}
\tag{B.85}
$$

$$\Pi_1^*(1,1) = \Pi_3^*(1,1) = \Big[4\mu\Big]^{-1}\Big[(\alpha+1)^2(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(\theta(16\theta(\theta+4)-313) \\ -1700)-3380)+492)+118548)+361248)-835848)-3923808) \\ +567632)+14604128)+9195776)-16299648)-14624256) \\ +2064384)+2737152)\Big], \tag{B.86}$$

$$\Pi_2^*(1,1) = -\mu^{-1}3(\alpha+1)^2(2\theta^2-21)(\theta(\theta(\theta(\theta(8\theta(\theta+2)-109)-196)+262) \\ +224)-56)^2, \tag{B.87}$$

$$\mu \equiv 2(-35\theta^6+717\theta^4-4226\theta^2+2184)^2. \tag{B.88}$$

Note we assume $\theta < \overline{\theta} \approx 0.145$ throughout this extension to allow well-defined solutions. This value is lower than in the main model as three firms compete in market B, and the highest value of externality must therefore be lower than when only firms 1 and 2 are competing. Indeed, the total output in market B with three firms is greater than with two firms (the relevant expressions are below, denoted by $Q^*_{B,2}, Q^*_{B,3}$, respectively).

$$Q^*_{B,2} = \Big[6(\theta^2-4)(6\theta^6-43\theta^4+88\theta^2-42)\Big]^{-1} \\ \Big[(\alpha+1)(\theta^2-3)(\theta(\theta(\theta(\theta(2\theta(4\theta-5)-89)+48)+258)-48)-144)\Big], \tag{B.89}$$

$$Q^*_{B,3} = \Big[70\theta^6-1434\theta^4+8452\theta^2-4368\Big]^{-1} \\ \Big[(\alpha+1)(\theta(\theta(\theta(\theta(\theta(11\theta-48)-390)+588)+3440)-672)-2016)\Big]. \tag{B.90}$$

Appendix C. Further extensions

In this appendix we (i) provide a microfoundation for the demand system of our main model; (ii) extend our analysis to incorporate consumer expectations over firms' quantity choices; and (iii) include same-side externalities in our analysis of price competition (from Section 5.1).

C.1. Microfoundation for demand

Here we present a microfoundation of the demand functions used in the main paper. We assume that consumers in one market benefit from data collected in the other market.[35]

Demand in market *A*. Consumers have a basic valuation r with support $[\alpha - \frac{\beta_A}{2}, \alpha + \frac{\beta_A}{2}]$ which follows the uniform distribution, i.e. $r \sim \mathcal{U}[\alpha - \frac{\beta_A}{2}, \alpha + \frac{\beta_A}{2}]$ where $\alpha \geq 0$ is the average willingness to pay and β is a measure of its volatility (heterogeneity).[36]

The monopolist firm in market A is able to increase the value of its product through two channels. First, the monopolist can incur a costly publicly observed investment which enhances consumers' value for the product. We denote the magnitude of this value enhancement as v_A which comes at an investment cost of $I(v_A)$. Second, the monopolist can collect data from market B which allows it to offer consumers a better product and increases the value of its product. Thus, the aggregate quality in market A depends also on the data collected (in our case proxied by demand) in market B, i.e., a larger demand served by the firm in market B allows it to collect more (or more representative) consumer data. Specifically, $q_1 + \Phi \cdot q_2$ is a proxy for the volume of data analyzed by firm 1. This value-enhancing data-driven network effect also depends on whether firm 1 has access to the data collected by its rival in market B. We denote by $\Phi \in \{0, 1\}$ the indicator function for data shared by firm 2 with firm 1's affiliate in market A, where $\Phi = 1$ indicates that firm 2 has shared data with firm 1 and $\Phi = 0$ indicates that firm 2 does not share data.

[35] We proxy demand in one market as a measure of data collected in that market.

[36] As in Katz and Shapiro (1985), the support also includes negative values, as some consumers may not find it worthwhile to purchase the product, even when the price is zero, unless the quality of the product exceeds a certain threshold. This specification also ensures that there is no corner solution and consumer demands are within the support when firms invest in value enhancement of their products.

We assume these two quality dimensions enter the utility expression in an additive and linear manner and thus we can write the utility expression of a consumer of type r in market A that consumes the product of firm 1 as

$$U_A(r) = r + v_A + q_1 + \Phi \cdot q_2 - P_A, \tag{C.1}$$

where v_A is the magnitude of value enhancement of firm $1'$s product in market A through investments, $q_1 + \Phi \cdot q_2$ is the level of value increase arising from data-driven network effects and P_A is the (implicit) price of the product.[37]

Consumers purchase from firm 1 in market A only when they obtain positive utility from doing so.[38] This condition pins down the demand for firm 1's product in market A as

$$U_A \geq 0 \implies r > \tilde{r}_A = P_A - v_A - (q_1 + \Phi \cdot q_2). \tag{C.2}$$

Thus, the mass of consumers purchasing from firm 1 in market A is

$$q_A = 1 - \frac{(\tilde{r}_A - \alpha + \frac{\beta_A}{2})}{\beta_A} = \frac{\alpha + \frac{\beta_A}{2} + v_A + q_1 + \Phi \cdot q_2 - P_A}{\beta_A}. \tag{C.3}$$

Inverting, we obtain the inverse demand function in market A, where $\mathcal{A} = \alpha + \frac{\beta_A}{2}$:

$$P_A(v_A, \Phi, q_A) = \mathcal{A} + v_A + q_1 + \Phi \cdot q_2 - \beta_A q_A. \tag{C.4}$$

Demand in market *B*. In market B where firm 1 competes with firm 2, we also assume that there are positive data externalities which enhance the value of the product sold by firm 1, for instance as the insights gained in market A are also useful in market B.[39]

[37] At the time consumers decide to purchase the product, they can observe participation by other consumers and take their consumption decision accordingly. Equivalently, firms could just announce their intended production quantities in the second stage, which they are then committed to, in the spirit of Katz & Shapiro (1985).

[38] Note that we assume that the outside option of consumers is zero. This assumption does not affect our results as any positive (and small enough) outside option is sufficient to provide qualitatively similar results.

[39] For instance, Google's collection of consumer data from the search market allows it to offer better services also in the smart speaker market. Another relevant example is Google's collection of traffic data in the maps market is expected to make its services more valuable to consumers in the self-driving car market.

In market B, consumers are heterogeneous in their basic valuation r with the support $[\gamma - \frac{\beta}{2}, \gamma + \frac{\beta_B}{2}]$ which follows the uniform distribution, i.e. $r \sim \mathcal{U}[\gamma - \frac{\beta_B}{2}, \gamma + \frac{\beta_B}{2}]$ where $\gamma \geq 0$ is the average willingness to pay and β is a measure of its volatility (heterogeneity). Further, as in market A competing firms in market B are also able to increase the value of the products for consumers through two channels. First, firms can invest in costly and public value-enhancing innovations that are observable to consumers. We denote this costly value creation by firm $i \in \{1, 2\}$ in market B as v_i. Secondly, if firms access data collected in market A, they can improve the value of their product in market B. To be more concrete, let's denote the positive data spillover of value generated by data from market A with demand q_A on the services of firm 1 as θq_A where q_A is a proxy for the insights gained from the data collected by firm 1 market A. Note that firm 2 has no presence in market A and does not benefit from this data advantage.

We naturally consider data generating cross-market value from market A to B and reciprocally, but our qualitative insights do not depend on this specification. The core mechanism of our analysis relies on the fact that the data generated by the small firm is useful for its larger competitor, which will in turn have incentives to accommodate it by softening the intensity of competition in market B. For instance, we would obtain similar results by considering a cross-market value of data only from market B to A.

Thus, the utility of a consumer of type r that buys from firms 1 and 2 is given as

$$U_1(r) = r + v_1 + \theta q_A - P_1, \quad U_2(r) = r + v_2 - P_1. \tag{C.5}$$

where v_i, θq_A and P_i for $i \in \{1, 2\}$ are respectively the quality levels invested by each firm, the consumer value for inter-market data driven value creation, and the price charged by each firm for its service in market B. Consumers buy the product of the firm that provides them the highest net utility. Under the above specification, firms 1 and 2 will have positive demand only if the quality-adjusted price at each firm $X = P_1 - v_1 - \theta q_A = P_2 - v_2$ is identical, implying that the "no arbitrage" condition holds.[40] Furthermore, we assume the value of a consumer's

[40]This condition implies that any consumer of type r should be indifferent between buying from firm 1 or 2 in market B i.e. $U_1(r) = U_2(r)$.

outside option is zero, so that consumers with $U_i < 0$ will not choose any firm. This implies that total demand is constituted only by those consumers for whom $r > X$. Hence, total demand in market B is $Q_B = 1 - \frac{(\Phi - \gamma + \frac{\beta_B}{2})}{\beta}$, where $Q_B = \sum_{i=1,2} q_i$ is the total output in market B. Rearranging and inverting the above total output for each firm i yields:

$$P_1(v_1, q_A, Q_B) = \gamma + \frac{\beta_B}{2} + v_1 + \theta q_A - \beta_B Q_B, \qquad P_2(v_2, 0, Q_B) = \gamma + \frac{\beta_B}{2} + v_2 - \beta_B Q_B. \quad \text{(C.6)}$$

Note that these demand functions are directly impacted by investments in value by each firm, i.e., each P_i is increasing in v_i. Additionally, in market B, because firms 1 and 2 are competing, their inverse demand function will also be (indirectly) impacted by the rival's investments through its strategic choice of the output which impacts the total output $Q_B = q_1 + q_2$.

Moreover, in favor of brevity and clear exposition of the results, we make the following variable transformation: $\mathcal{B} = \gamma + \frac{\beta_B}{2}$.

C.2. Consumer expectations

In markets with demand externalities, a consumer's willingness to pay is determined by how many other consumers buy. In our one-shot model, this is implemented in a straightforward manner: firms choose their production quantities in all markets during the last stage, which consumers observe.[41] It follows that consumers are willing to buy as dictated by the demand curves specified at those sales levels.[42] An alternate approach is to model consumers' expectations on other consumers' purchase decisions. We offer such a version of our main model in this extension. Specifically, when consumers decide to buy in one market they have an expectation of how much (quality-enhancing) data is generated in the other market (rather than knowing it, as in our main model). The introduction of output expectations reduces the sensitivity of profits to output (because consumers do not react to a deviation in output from the expected level)

[41] Equivalently, firms commit to production quantities (Katz & Shapiro 1985, pp439–440).

[42] Another way to interpret the game is to consider that the (unique) market-clearing price that follows from the chosen production quantities prevails in the market. Consumers, looking at the market, can then simply consult their own willingness to pay and compare it against that price. Those with the highest willingness to pay (higher than the price when no-one else buys) buy first, which pushes up other demand curves, encouraging more to buy, and so on. The purchasing ends once all the produced units have been purchased.

and so we obtain different quantitative predictions from our main model, but the mechanism and qualitative results are unchanged.

Denote by q_A^e the expected output in market A and by $q^e(q_1^e, q_2^e, \Phi) = q_1^e + \Phi \cdot q_2^e$ the expected level of quality-enhancing data generated in market B. The (inverse) demand functions are:

$$P_A(v_A, q^e(\cdot), q_A^e) = \mathcal{A} + v_A + \theta q^e(\cdot) - \beta_A q_A, \tag{C.7}$$

$$P_1(v_1, q_A^e(\cdot), Q_B) = \mathcal{B} + v_1 + \theta q_A^e(\cdot) - \beta_B Q_B, \tag{C.8}$$

$$P_2(v_2, Q_B) = \mathcal{B} + v_2 - \beta_B Q_B. \tag{C.9}$$

As in Katz & Shapiro (1985), our solution concept is Fulfilled Expectations Cournot Equilibrium in which each firm chooses its output given consumers' expectations, and in equilibrium those expectations are required to be correct. Including expectations in the model alters the relationship between output and profits. A firm's output choice in one market no longer directly influences demand in the other. Instead, consumers' expectations of that output choice do, i.e.,

$$\frac{\partial P_A(\cdot)}{\partial q_1} = \frac{\partial P_A(\cdot)}{\partial q_2} = \frac{\partial P_1(\cdot)}{\partial q_A} = \frac{\partial P_2(\cdot)}{\partial q_A} = 0, \tag{C.10}$$

and so the first order conditions for firm 1 are now respectively:

$$\frac{\partial \Pi_1}{\partial q_A} = P_A(\cdot) + \frac{\partial P_A(\cdot)}{\partial q_A} q_A \quad \text{and} \quad \frac{\partial \Pi_1}{\partial q_1} = P_1(\cdot) + \frac{\partial P_1(\cdot)}{\partial Q_B} q_1, \tag{C.11}$$

which one can show to be below those of our main model. The reduced responsiveness of own profits to own output affects equilibrium output and investment levels, but these quantitative differences do not impact the qualitative results and firm 2 is willing to share its data for free.

C.3. Price competition with same-side data externalities

Here we extend the model in Section 5.1 to show that the mechanism we identified holds when firms are differentiated, compete in prices, and when data generates both cross and within-market externalities in market B.

We continue with the notation of Section 5.1 and similarly assume consumers' outside options are given by $s \sim \mathcal{U}[0,1]$ so that the valuations are described by (11). Consumers buy when $s < \tilde{s}(v_A, p_A, \Phi)$, and so demand in market A is $q_A(v_A, p_A, \Phi) = \tilde{s}(\cdot)$. In market B, the generalist firm 1 and the incumbent specialist firm 2 compete à la Hotelling, and the utility of a consumer of type x from firm 1 and 2's products are, respectively,

$$u_1(v_1, p_1, x) = v_1 + \theta\left(q_A^e + q_1^e + \Phi q_2^e\right) - p_1 - tx, \tag{C.12}$$

$$u_2(v_2, p_2, x) = v_2 + \theta q_2^e - p_2 - t(1-x), \tag{C.13}$$

where q_A^e, q_1^e, and q_2^e are the expected values derived from the data collected in market A and by firm 1 and firm 2 in market B, respectively. Using the same reasoning as before, we denote the indifferent consumer in market B by $\tilde{x}$, characterized by $u_1(\cdot) = u_2(\cdot) \implies \tilde{x}(v_1, v_2, p_1, p_2, q_A^e)$. Using this expression we can characterize the demands:

$$q_1(v_1, v_2, p_1, p_2, q_A^e) = \tilde{x}(\cdot), \quad q_2(v_2, v_1, p_2, p_1, q_A^e) = 1 - \tilde{x}(\cdot). \tag{C.14}$$

Profits and the timing of the game are otherwise as in Section 5.1. We make the following technical assumptions to ensure that the second order conditions are satisfied and that we obtain an interior solution.

Assumption 2. *(i) The cross-network data externality θ is not too strong, $\theta < \tilde{\theta} \approx 0.137$; (ii) the transportation cost parameter is such that $\frac{15\theta^2+18\theta+2}{18} < t < \frac{-9\theta^2+30\theta+8+\sqrt{225\theta^4+180\theta^3-108\theta^2-96\theta+64}}{72}$.*

To solve the game, we first consider the case in which firm 2 does not share data.

Specialist firm does not share data

At stage 3, given expectations consumers act and their demands are

$$q_1(v_1, v_2, p_1, p_2, q_1^e, q_A^e) = \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2) + (q_A^e + q_1^e - q_2^e)}{2t}, \tag{C.15}$$

$$q_2(v_2, v_1, p_2, p_1, q_2^e, q_A^e) = 1 - q_1(\cdot), \tag{C.16}$$

$$q_A(v_A, p_A, q_1^e) = v_A + \theta q_1^e - p_A. \tag{C.17}$$

In equilibrium, consumers' expectations should match the outcome and we set $q_A^e = q_A$, $q_1^e = q_1$, $q_2^e = q_2$ and solve for demands to get demands as a function of price and investment levels.

$$q_1(v_1, v_2, v_A, p_1, p_2, p_A) = \frac{t - p_1 + p_2 + \theta(v_A - p_A - 1) + v_1 - v_2}{2t - \theta(2+\theta)}, \tag{C.18}$$

$$q_2(v_2, v_1, v_A, p_2, p_1, p_A) = 1 - q_1(\cdot), \tag{C.19}$$

$$q_A(v_A, v_1, v_2, p_A, p_1, p_2) = \frac{t(2v_A + \theta) - \theta(2v_A + v_2 + \theta + p_1 - p_2 - v_1) + 2p_A(\theta - t)}{2t - \theta(2+\theta)}. \tag{C.20}$$

Substituting these demands in the profit expression yields

$$\Pi_1 = \underbrace{p_A q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1 q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2 q_2(\cdot) - I(v_2). \tag{C.21}$$

Differentiating the profit of the generalist firm with respect to its prices p_1 and p_A and the specialist firm with respect to p_2 gives a system of first order conditions. Solving those yields prices as functions of investment levels given as follows.

$$p_1(v_1, v_2, v_A) = t + \frac{2v_1 - 2v_2 - \theta(2\theta + v_A + 6)}{6}, \tag{C.22}$$

$$p_2(v_2, v_1, v_A) = t + \frac{v_2 - v_1 - \theta(2\theta + v_A + 3)}{3}, \tag{C.23}$$

$$p_A(v_A, v_1, v_2) = \frac{v_A}{2}. \tag{C.24}$$

Substituting these prices into the profit expression yields profits as functions of investments.

$$\Pi_1 = \underbrace{p_A(\cdot)q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1(\cdot)q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2(\cdot)q_2(\cdot) - I(v_2). \tag{C.25}$$

Differentiating Π_1 with respect to its prices, v_1 and v_A and Π_2 with respect to v_2, and solving the resulting system of first order conditions yields

$$v_1^\star(0) = \frac{18t - 4 - 6\theta(\theta+3)}{54t - 3(\theta(13\theta+18)+4)}, \tag{C.26}$$
$$v_2^\star(0) = \frac{20\theta^2 + 18\theta - 18t + 4}{39\theta^2 + 54\theta - 54t + 12}, \tag{C.27}$$
$$v_A^\star(0) = -\frac{4\theta(3\theta(\theta+3) - 9t + 2)}{54t - 3(\theta(13\theta+18)+4)}. \tag{C.28}$$

Substituting these investment output into prices, demand expression and profit yields

$$p_1^\star(0) = v_1^\star(0)\left(3t - \frac{\theta(6+5\theta)}{2}\right),\ p_2^\star(0) = 3v_2^\star(0)\left(t - \frac{\theta(2+\theta)}{2}\right),\ p_A^\star(0) = \frac{v_A^\star(0)}{2}, \tag{C.29}$$
$$q_1^\star(0) = \frac{3v_1^\star(0)}{2},\ q_2^\star(0) = \frac{3v_2^\star(0)}{2},\ q_A^\star(0) = \frac{5v_A^\star(0)}{4}, \tag{C.30}$$
$$\Pi_1^\star(0) = \frac{v_A^\star(0)^2}{8} + \frac{v_1^\star(0)^2}{2}\left(9t - \frac{3\theta(5\theta+2)}{2} - 1\right), \tag{C.31}$$
$$\Pi_2^\star(0) = \frac{v_2^\star(0)^2}{2}\left(9t - \frac{9\theta(2+\theta)}{2} - 1\right). \tag{C.32}$$

Specialist firm shares data

At stage 3, given expectations, consumers act and their demands are

$$q_1(v_1, v_2, p_1, p_2, q_1^e, q_A^e) = \frac{1}{2} + \frac{v_1 - v_2 - (p_1 - p_2) + (q_A^e + q_1^e)}{2t}, \tag{C.33}$$
$$q_2(v_2, v_1, p_2, p_1, q_2^e, q_A^e) = 1 - q_1(\cdot), \tag{C.34}$$
$$q_A(v_A, p_A, q_1^e, q_2^e) = v_A + \theta(q_1^e + q_2^e) - p_A. \tag{C.35}$$

In any equilibrium, consumers' expectations should match the outcome and so we set $q_A^e = q_A$, $q_1^e = q_1$ and $q_2^e = q_2$ and solve for demands as functions of price and investment levels.

$$q_1(v_1, v_2, v_A, p_1, p_2, p_A) = \frac{t + v_1 - v_2 + p_2 - p_1 + \theta(\theta - p_A + v_A)}{2t - \theta}, \tag{C.36}$$

$$q_2(v_2, v_1, v_A, p_2, p_1, p_A) = 1 - q_1(\cdot), \tag{C.37}$$

$$q_A(v_A, p_A) = \theta + v_A - p_A. \tag{C.38}$$

Substituting these demands in the profit expression yields

$$\Pi_1 = \underbrace{p_A q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1 q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2 q_2(\cdot) - I(v_2). \tag{C.39}$$

Differentiating the profit of the generalist firm with respect to its prices p_1 and p_A and the specialist firm with respect to p_2 gives a system of first order conditions. Solving them yields prices as functions of investment levels given as follows.

$$p_1(v_1, v_2, v_A) = \frac{(2t - \theta)(6t + 2v_1 - 2v_2 + \theta(\theta + v_A - 2))}{12t - \theta(\theta + 6)}, \tag{C.40}$$

$$p_2(v_2, v_1, v_A) = \frac{(2t - \theta)(6t - 2v_1 + 2v_2 - \theta(2\theta + v_A + 4))}{12t - \theta(\theta + 6)}, \tag{C.41}$$

$$p_A(v_A, v_1, v_2) = \frac{3t(\theta + 2v_A) - \theta(\theta(\theta + 2) + v_1 - v_2 + (\theta + 3)v_A)}{12t - \theta(\theta + 6)}. \tag{C.42}$$

Substituting these prices into the profit expression yields profits as functions of investments.

$$\Pi_1 = \underbrace{p_A(\cdot) q_A(\cdot) - I(v_A)}_{\text{Market A profit}} + \underbrace{p_1(\cdot) q_1(\cdot) - I(v_1)}_{\text{Market B profit}}; \qquad \Pi_2 = p_2(\cdot) q_2(\cdot) - I(v_2). \tag{C.43}$$

Differentiating Π_1 with respect to its prices, v_1 and v_A and Π_2 with respect to v_2, and solving the resulting system of first order conditions yields

$$v_1^\star(1) = \mathcal{C}^{-1}(8t - \theta(\theta+4))\left(-\theta((\theta-1)\theta(\theta+6)-4) + 36t^2 + \left(9\theta^2 - 30\theta - 8\right)t\right), \tag{C.44}$$

$$v_2^\star(1) = \mathcal{C}^{-1}4(2t-\theta)\left(\theta(\theta(\theta(2\theta+13)+13)+4) + 36t^2 - (\theta(25\theta+42)+8)t\right), \tag{C.45}$$

$$\begin{aligned} v_A^\star(1) = \mathcal{C}^{-1}\theta\Big(& -\theta^2(\theta(\theta(5\theta+53)+150)+64) + 1152t^3 \\ & - 4\left(57\theta^2 + 420\theta + 64\right)t^2 + \theta(\theta(\theta(11\theta+218)+852)+256)t\Big). \end{aligned} \tag{C.46}$$

$$\mathcal{C} \equiv (12t - \theta(\theta+6))\left(\theta(\theta(\theta(\theta+8)+19)+8) + 72t^2 - 8(\theta(2\theta+9)+2)t\right) \tag{C.47}$$

Substituting these investment output into prices, demand expression and profit yields

$$p_1^\star(1) = v_1^\star(1)\frac{(2t-\theta)(12t-\theta(6+\theta))}{8t-\theta(4+\theta)},\ p_2^\star(1) = v_2^\star(1)\left(3t - \frac{\theta(6+\theta)}{4}\right), \tag{C.48}$$

$$p_A^\star(1) = \theta\frac{\begin{pmatrix} 792t^3 - 2(3\theta(47\theta+200)+88)t^2 + \theta(\theta(\theta(30\theta+283)+634)+176)t \\ -\theta^2(\theta(\theta(\theta(\theta+15)+72)+116)+44) \end{pmatrix}}{\mathcal{C}}, \tag{C.49}$$

$$q_1^\star(1) = v_1^\star(1)\left(\frac{(12t-\theta(6+\theta))}{8t-\theta(4+\theta)}\right),\ q_2^\star(1) = v_2^\star(1)\left(\frac{(12t-\theta(6+\theta))}{8t-4\theta}\right), \tag{C.50}$$

$$q_A^\star(1) = \frac{\begin{pmatrix} 1224t^3 - 2(3\theta(35\theta+296)+136)t^2 \\ \theta(\theta(\theta(9\theta+199)+894)+272)t - 4\theta^2(\theta(\theta(\theta+12)+39)+17) \end{pmatrix}}{\mathcal{C}}, \tag{C.51}$$

$$\Pi_1^\star(1) = p_A^\star(1)q_A^\star(1) - \frac{(v_A^\star(1))^2}{2} + \frac{(v_1^\star(1))^2}{2}\left(\frac{(2t-\theta)(12t-\theta(6+\theta))^2}{(8t-\theta(4+\theta))^2} - 1\right), \tag{C.52}$$

$$\Pi_2^\star(1) = \frac{(v_2^\star(1))^2}{2}\left(\frac{(12t-\theta(6+\theta))^2}{8(2t-\theta)} - 1\right). \tag{C.53}$$

Comparing firm 2's profit under data sharing with the case when firm 2 does not share yields:

$$\Pi_2^\star(1) - \Pi_2^\star(0) > 0. \tag{C.54}$$

Firm 2 profits by sharing data with firm 1 (given our regularity condition Assumption 2).