Data Sharing and Collusion on Hybrid Platforms

Kevin Mi*

July 7, 2025

Abstract

Many online retail platforms, such as Amazon and JD.com, have recently begun providing sellers with proprietary consumer data. In this paper, we investigate how the sharing of such data can incentivize seller collusion through personalized pricing. We find that the effect of data sharing on collusion sustainability and profitability depends on the mode operated by the platform. When a platform acts as both a host and seller, sharing data leads to more collusive outcomes. However, when a platform only intermediates purchases, sharing data hinders collusion. Our results suggest that imposing a ban on data sharing may be ineffective and harmful to consumer welfare.

Keywords: data sharing, collusion, platform, personalized pricing

JEL Codes: D21, D40, L11

*Department of Economics, Michigan State University, E-mail: kwmi@msu.edu

1 Introduction

Online platforms, such as Amazon and JD.com, serve as intermediaries between buyers and sellers. This "gatekeeper" position allows them to collect large amounts of consumer-level data through tracking cookies, registration information, loyalty programs, etc. Moreover, platforms have begun sharing this proprietary consumer data with their sellers. For example, Amazon shares buyer identification and transaction information with their third-party sellers (see Choe et al., 2024). JD.com has also created a service that provides third-party sellers with consumer-level information that includes credit history, location, search history, and other preferences[1]. Both such acts assist sellers in offering targeted coupons to buyers or setting different prices based on consumer characteristics.

The implications of personalized pricing on consumer and seller welfare have been widely discussed in recent years by both policymakers and researchers. For instance, OECD (2018) states that personalized pricing "is typically pro-competitive and often enhances consumer welfare," but may also be harmful "by potentially enabling the exploitation of consumers and creating a perception of unfairness." The ambiguity in results can be somewhat seen in Rhodes and Zhou (2024). They find that the effect of personalized pricing on consumer welfare depends on whether full market coverage is satisfied. Although the static effects of personalized pricing are well understood, the effect that sharing data has on collusive outcomes has not been heavily explored.

Moreover, an increasing number of decisions are now being supported through machine learning and algorithms that leverage big data. Indeed, pricing algorithms are being more widely adopted by sellers and platforms can even provide sellers with simple pricing tools. Amazon provides a repricing software that is designed to automate pricing for sellers.[2] Previous research has shown that the usage of rule-based pricing strategies on Amazon has become more prevalent (see Chen et al., 2016, and Wang et al., 2022).

The emergence of such advanced tools has posed two major difficulties to fair competition on these online marketplaces. First, policymakers have recently expressed concern about how sharing proprietary consumer data combined with the use of digital pricing tools potentially contributes price discrimination. For example, the UK's Competition and Markets Authority (2018) warns that pricing algorithms "in combination with the growth of 'Big Data'...might lead to personalized pricing." Researchers have also found that algorithms can lead to raising prices and even learn to autonomously price collude without any prior knowledge of the market environment. One notable example of price hikes through algorithms is Amazon's

[1]Hu et al. (2023) provides a more detailed description of how JD.com shares consumer information.

[2]See `https://aws.amazon.com/marketplace/pp/prodview-lsel7dks2e7uw`.

scrapped "Project Nessie" which resulted in a lawsuit filed by the FTC.[3] Also, the German Monopolies Commission (2018) states that "the increasing use of pricing algorithms makes collusion-related consumer damages more likely in the future." In addition, the question of whether antitrust liability can be established when various decisions are made by humans versus machines has been widely discussed (see, eg., OECD, 2017). Thus, the introduction and increasing use of pricing algorithms poses a deep concern for competition regulators due to its ability to price discriminate and set supra-competitive prices.

One policy remedy to address these issues is the *Preventing Algorithmic Collusion Act of 2024* which was introduced by the U.S. Congress that prohibits the use of a pricing algorithm that incorporates or was trained using proprietary data.[4] Although the legislation provides a first-step solution towards preventing algorithmic collusion, does restricting data supplied to seller, assuming they are using pricing algorithms, really reduce collusive sustainability?[5] How would this depend on the design of the marketplace or the platform's incentive to share data?

In this paper, we analyze how sharing data affects sustainability of tacit collusion between two horizontally differentiated sellers intermediated by a data-rich platform.[6] We consider two different modes operated by the platform. First, we consider a platform that runs in "marketplace mode" where the platform's role is providing access between buyers and sellers. We also consider when the platform runs in "dual mode" where the platform is fully integrated with a seller. Under dual mode, the platform acts as both host and seller, e.g. Amazon with their Amazon Basics products. This means the platform sets prices in competition with the other seller.

We develop an infinitely repeated game where, given the market structure of the platform, the stage game is as follows. In each time period, the data-rich platform sets an ad-valorem fee and chooses whether to share its data with sellers. If sellers have access to data, they can set personalized prices for each consumer instead of a standard uniform price. Observing their prices, consumers make their purchasing decisions. We assume each strategic player follows grim-trigger strategies to investigate their incentives for collusion.

[3]For experimental evidence of algorithmic collusion, see Calvano et al. (2020) and Klein (2021). Note that there is also some empirical evidence that suggests pricing algorithms may benefit consumers, see Hanspach et al. (2024).

[4]See `https://www.congress.gov/bill/118th-congress/senate-bill/3686`.

[5]Others have addressed the remedy proposed in the *Preventing Algorithmic Collusion Act* through different contexts. For example, Harrington Jr (2025) discusses the legislation in a hub (algorithm distributor) and spoke (firms) context and finds that banning the use of proprietary data for pricing algorithms may be ineffective or create additional inefficiencies.

[6]We treat the mode of the platform as exogenous. Practically, platforms may choose which mode to operate based on the market environment. However, we are primarily focused on potential data sharing implications instead.

Our first main result explores collusive outcomes given the platform operates as a marketplace. In this scenario, sellers face no information asymmetry. Under competition, platforms extract profits from the sellers to the point that they are on the margin of staying in the marketplace regardless of data sharing policy. Without data, price collusion is less profitable for sellers. But as sellers must internalize some profit loss on their turf when setting a lower uniform price, sellers are less incentivized to deviate. That is, collusion is more likely when under marketplace mode. Collusive profitability, from a platform perspective, depends on how patient sellers are. Because collusion is less profitable for sellers, platforms must set lower fees to ensure collusion. Thus, platforms generally prefer sharing data if sellers are sufficiently patient, vice versa.

If the platform operates under dual mode, sharing data in fact increases collusive sustainability and profitability. Now, the data-rich platform competes with the seller and has an informational advantage under no data sharing. The same logic applies as before, but with a reduced effect. That is, even though collusion is less profitable for the seller under no data sharing versus full data sharing, the seller must set a lower uniform price when deviating and incur a loss from consumers originally in their turf. However, because the platform has an advantage under collusion, this loss is greatly reduced in comparison to when the platform was in marketplace mode. This is the driving force behind the sustainability result. When both players have access to consumer data, collusive profits are highest and the platform can enforce a higher fee under collusion. Thus, profitability for the platform is greater with data sharing. Table 1 summarizes the key findings in the paper.

As consumer welfare is lowest when sellers price collude and have access to data, our analysis shows that the effects of data sharing on collusive outcomes are not black and white. From a regulation standpoint, banning the usage of proprietary data for pricing algorithms may be ineffective. Instead, regulation policies should depend on the structure of the platform or specific product markets. These results are also robust when considering more flexible levels of data sharing, privacy concerns form consumers, and harsher punishment strategies from sellers.

The rest of this paper goes as follows. Section 2 reviews the related literature. Section 3 describes the model of the stage game and dynamic strategy of the players. Sections 4 and 5 analyze collusive outcomes under marketplace mode and dual mode, respectively. Section 6 provides several extensions such as incorporating more flexible levels of data sharing, privacy concerns from consumers, and harsher punishment strategies from sellers. Section 7 ends with concluding remarks.

	Collusive Sustainability	**Collusive Profitability**
Marketplace Mode	*No Data Sharing*	*No Data Sharing − if sellers are less patient* *Data Sharing − if sellers are more patient*
Dual Mode	*Data Sharing*	*Data Sharing*

Table 1: Platform preference on which mode to operate is based on collusive sustainability or profitability.

2 Related literature

This paper contributes to the literature on competitive personalized pricing. Our stage game follows a framework similar to that of Thisse and Vives (1988) which has been widely adopted to study spatial price discrimination. They find that personalized pricing reduces firm profits and increases consumer surplus under competition. Other related papers that use this framework include Shaffer and Zhang (2002) and Montes et al. (2019). Shaffer and Zhang (2002) studies competition between firms with asymmetric targeting costs for personalized pricing, whereas Montes et al. (2019) examines how a data intermediary should distribute data to firms for the purpose of personalized pricing. Anderson et al. (2023) and Rhodes and Zhou (2024) also focus on competitive personalized pricing by using a general discrete-choice model. Specifically, Anderson et al. (2023) allows consumers to opt-in to discounts which is costly to firms, while Rhodes and Zhou (2024) show that the welfare consequences discussed in Thisse and Vives (1988) can be reversed if the market is not fully covered.

We also touch on the viability of data sharing among competing sellers. Gradwohl and Tennenholtz (2023) find that selling only some consumers' data may lead to win-win outcomes for competing sellers. Both Hu et al. (2023) and Navarra et al. (2024) model a dual-role platform that can share data with its seller that it hosts. In particular, Hu et al. (2023) find that information sharing between the platform and seller should be regulated based on commission rate. Navarra et al. (2024) show that platforms are always incnetivized to share data, albeit they consider other data sharing scenarios than just full data sharing. We instead find that platforms unambiguously benefit by data sharing only under dual mode. Our analysis mainly differs in that we consider ad-valorem fees as opposed to transaction fees we are primarily focused on collusive outcomes under such information sharing policies.

At its core, our paper contributes to the literature that examines the interplay between collusion and price discrimination. Specifically, previous literature has shown that the effect price discrimination has on collusive behavior is ambiguous. For example, Helfrich and

Herweg (2016) use a linear city model and finds that a ban on third-degree price discrimination increases collusion sustainability. Döpper and Rasch (2024) similarly shows that a ban on second-degree price discrimination facilitates collusive outcomes. However, Peiseler et al. (2022) analyzes a setting where firms can third-degree price discriminate under private information. The addition of noisy signals on consumer preferences yields the opposite outcome. That is, they show that when signals are sufficiently noisy, a ban on third-degree price discrimination may reduce collusion. To the best of our knowledge, we are the first to bridge personalized pricing using a Hotelling framework and collusion.

Moreover, we address questions brought up in the fierce debate on algorithmic collusion. Many seminal works have discussed how pricing algorithms may autonomously collude. For example, Calvano et al. (2020) and Klein (2021) simulate competing reinforcement-learning algorithms and find that they can converge to collusive outcomes. In contrast, Miklós-Thal and Tucker (2019) suggest that better algorithms and demand forecasting potentially leads to lower prices and higher consumer welfare. Johnson et al. (2023) tackles the question from a platform design perspective. They show that a platform may enforce steering rules that can combat algorithmic collusion to the benefit of both consumers and the platform.

3 Description of the model

Consider the following infinitely repeated game that represents a specific product market on a data-rich platform. We first describe the model when the platform operates in marketplace mode and further discuss how we adapt the model when platform also operates as a seller in Section 5.

There are two competing sellers B_1 and B_2 hosted by the platform. Specifically, sellers B_1 and B_2 can produce goods at marginal cost 0 and are positioned at the beginning and end of a Hotelling (1929) line, respectively (i.e. $l_{B_1} = 0$ and $l_{B_2} = 1$). A measure 1 of consumers are uniformly distributed on the Hotelling line. Consumers have at most one unit of demand, value both sellers' good at 1, and have an outside utility of 0. Also, consumers face a transportation cost τ per unit traveled where $0 < \tau < 2/7$.[7] We interpret τ as the degree of product differentiation. To summarize, a consumer located at $x \in [0, 1]$ purchasing a good from $i \in \{B_1, B_2\}$ gains the following utility:

$$U(x, p_i) = 1 - p_i - \tau|l_i - x|$$

[7]The restrictions on τ guarantee there will be full market coverage if a seller prices as a monopolist which is key in our analysis. Also, this restriction significantly simplifies the analysis for deviating outcomes, but is not needed for this purpose. We can in fact have $\tau < \frac{1}{2}$ and still yield the same results.

where p_i is the price set by i.

We assume that the platform is data-rich in the sense that it knows all consumer locations along the unit interval. Moreover, platform A has the option to share this consumer-level information with the sellers via its data-sharing policy $\mathcal{D} \in \{S, N\}$. That is, if $\mathcal{D} = S$, the platform shares data on every consumer with the sellers, allowing sellers to enforce personalized pricing (i.e. first-degree price discrimination). If $\mathcal{D} = N$, the sellers sets a uniform price instead. In each period, the platform A posts an ad-valorem fee $r \in [0, 1]$ and its data-sharing policy $\mathcal{D}$. The sellers observe the posted r and $\mathcal{D}$ and decide whether to join the platform before making their pricing decisions.

We require two crucial assumptions in our environment. First, each seller has a reservation payoff of u where $0 < u < \frac{\tau}{8}$. This can be a result of the seller having a direct channel outside the platform or selling on some competing platform. Without a nonnegative outside option, we would have trivial results, as the platform will always set $r = 1$ leaving sellers with no profits. The other crucial assumption is that the platform always ensures sellers stay on the platform by setting their fee such that seller profits are weakly greater than u. Specifically, we prefer to focus on the competitive and collusive outcomes based on different data-sharing policies without the distraction of our sellers' choice of entry and exit from the platform. Of course, there may be many motives for the platform's preference to keep their seller. We provide one possible explanation by extending our base model in Appendix B where all the results presented in Sections 4 and 5 still hold.

The timing of the stage game goes as follows:

1. A chooses $(r, \mathcal{D}) \in [0, 1] \times \{S, N\}$.

2. Sellers observe $(r, \mathcal{D})$ and set personalized prices or uniform prices depending on $\mathcal{D}$. All pricing is done simultaneously.

3. Consumers observe prices, make their purchasing decisions, and outcomes are realized.

3.1 Dynamic Strategy

We assume that all players are long-lived and share a common discount factor $\delta \in (0, 1)$ per period. Furthermore, under perfect monitoring, B_1 and B_2 follow grim trigger strategies *a la* Friedman (1971). That is, if any seller deviates from collusive pricing, the other seller prices according to the (static) Nash equilibrium of the stage game for each subsequent period. The platform's objective is to maximize its joint discounted profit. We stick with sellers following grim-trigger strategies for the following reasons. First, in the context of Hotelling set ups, there is evidence that optimal punishment strategies tend to be similar to grim-trigger strategies (see Häckner (1996) and Liu and Serfes (2007)). Recent studies on algorithmic collusion have also shown that algorithms can learn sophisticated grim-trigger strategies to enforce supra-competitive prices. Finally, optimal punishment strategies come with the trade-off of less tractable models. We will provide more specific details on how we characterize collusive sustainability in Sections 4 and 5.

4 Equilibrium in the non-integrated case

We first proceed by solving for the Subgame Perfect Nash Equilibrium (SPNE) as well as collusive and deviation outcomes in the stage game under two benchmark cases: full data sharing where $\mathcal{D} = S$ and no data sharing where $\mathcal{D} = N$. Because sellers are symmetric and face the same fee, their profits under competition, collusion, and when deviating are the exact same.

4.1 Full Data Sharing, $\mathcal{D} = S$

Suppose A sets some fee r and there are two sellers on the platform. The competitive pricing follows exactly from Thisse and Vives (1988). Seller B_1 sets price $\max\{\tau(1-2x), 0\}$ and seller B_2 sets price $\max\{\tau(2x-1), 0\}$ for a consumer located at x. As a result, seller payoffs can be characterized as

$$\pi_S^*(r) = (1-r)\int_0^{\frac{1}{2}} \tau(1-2x)dx = (1-r)\frac{\tau}{4}. \tag{1}$$

By backwards induction, platform A will set r to extract seller profit up to u. Specifically, A's optimal ad valorem fee r_S^* must satisfy $\pi_S^*(r_S^*) = u \iff r_S^* = 1 - \frac{4u}{\tau}$. Thus, under competition, A gains $\pi_{A,S} = 2\pi_S^*(r_S^*)\frac{r_S^*}{1-r_S^*} = \frac{\tau}{2} - 2u$ and both sellers end with u in profit. Under data sharing, each seller has full knowledge of consumer locations and tries to win over consumers on its rival's side with low prices. This induces sellers to charge low prices even for consumers who have strongly prefer their product. Naturally, the platform will then calibrate their fee to extract as much profit as possible.

Under full collusion, sellers can maximize industry profits by splitting the market and setting their price to each consumer's willingness to pay. This occurs with the following price schedule

$$p_S^c(x) = \begin{cases} 1-\tau x & \text{if } x \leq \frac{1}{2} \\ 1-(1-x)\tau & \text{if } x \geq \frac{1}{2} \end{cases}. \tag{2}$$

That is, each seller would achieve a collusive profit of

$$\pi_S^c(r) = (1-r)\int_0^{\frac{1}{2}}(1-\tau x)dx = (1-r)(\frac{1}{2}-\frac{\tau}{8}). \tag{3}$$

Now given that a seller sets $p_S^c(x)$, the opposing seller's optimal deviating strategy would be to undercut $p_S^c(x)$ for all consumers. B_1's deviating price would then be $1-\tau x$ and B_2's deviating price would then be $1-\tau(1-x)$ for consumer located at x. Thus, deviating profit under full data sharing is

$$\pi_S^d(r) = (1-r)\int_0^{1}(1-\tau x)dx = (1-r)(1-\frac{\tau}{2}). \tag{4}$$

4.2 No Data Sharing, $\mathcal{D} = N$

Again, suppose A sets some fee r. Similarly to the full data sharing scenario, the competitive payoffs follow closely from Thisse and Vives (1988) where each seller sets uniform price τ and obtains

$$\pi_N^*(r) = (1-r)\frac{\tau}{2}. \tag{5}$$

Using the same logic from before, A's optimal fee r_N^* is such that $\pi_N^*(r_N^*) = u \iff r_N^* = 1-\frac{2u}{\tau}$. Thus, A will attain $\pi_{A,N} = 2\pi_N^*(r_N^*)\frac{r_N^*}{1-r_N^*} = \tau - 2u$ and both sellers are left with u profit. Here, we get the Thisse and Vives (1988) result where personalized pricing benefits every consumer. But, this also implies that the platform prefers to restrict data sharing in competition as their fee is dependent on seller pricing, i.e. $r_N^* > r_S^*$.

Because sellers are constrained to setting uniform prices, both sellers will set $p_N^c = 1-\frac{\tau}{2}$ under full collusion. Like Full Data Sharing, sellers will split the market and each consumer purchases from the seller that is closer to them. However, as the sellers have no knowledge of individual locations, they cannot fully extract the consumer surplus. If a seller deviates, they would set their uniform price at $p_N^d = 1-\frac{3\tau}{2}$ to capture the entire market.[8] Thus, the collusive and deviating profits under no data sharing can be characterized as follows:

$$\pi_N^c(r) = (1-r)(\frac{1}{2}-\frac{\tau}{4}), \tag{6}$$

$$\pi_N^d(r) = (1-r)(1-\frac{3\tau}{2}). \tag{7}$$

[8] This is not necessarily optimal if $\frac{2}{7} \leq \tau < \frac{1}{2}$. In such a case, the seller would only undercut and gain a portion of the market. Fortunately, this does not change our main results and our insights remain consistent.

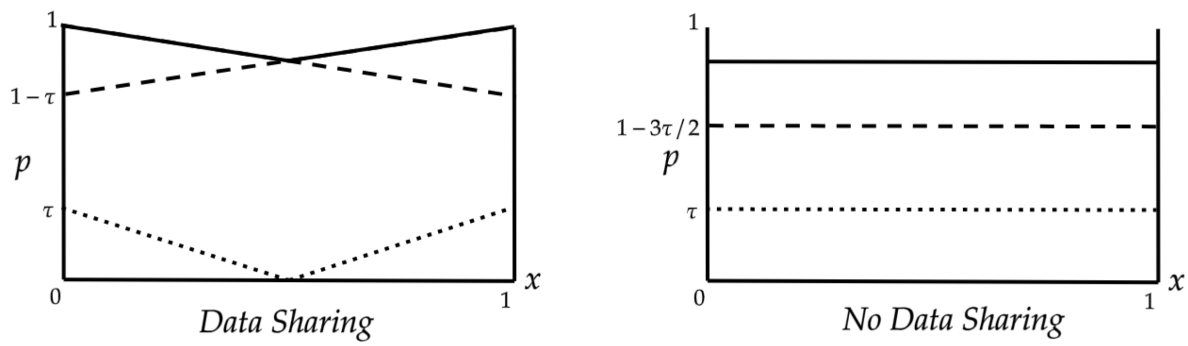

Figure 1: A graphical comparison of the price schedules based on different pricing policies under marketplace mode. Solid lines represent collusive prices, dashed lines represent deviating prices, and dotted lines represent competitive prices.

4.3 Comparison of Full Data Sharing vs. No Data Sharing

Given the dynamic environment presented, we say that collusion is sustainable under policy $\mathcal{D} \in \{S, N\}$ as an SPNE if and only if $\frac{\pi^c_{\mathcal{D}}(r)}{1-\delta} \geq \pi^d_{\mathcal{D}}(r) + \frac{\delta}{1-\delta}\pi^*_{\mathcal{D}}(r^*_{\mathcal{D}})$ and $2\pi^c_{\mathcal{D}}(r)\frac{r}{1-r} \geq \pi_{A,\mathcal{D}}$ where r is the ad valorem fee set under collusive pricing for policy $\mathcal{D}$. This means δ must satisfy

$$\delta \geq \delta^{mkt}_{\mathcal{D}}(r) \equiv \frac{\pi^d_{\mathcal{D}}(r) - \pi^c_{\mathcal{D}}(r)}{\pi^d_{\mathcal{D}}(r) - \pi^*_{\mathcal{D}}(r^*_{\mathcal{D}})} \text{ for } \mathcal{D} \in \{S, N\}.$$

Here, $\delta^{mkt}_{\mathcal{D}}(r)$ represents the critical discount factor under benchmark policy $\mathcal{D}$ and measures collusive sustainability. A higher $\delta^{mkt}_{\mathcal{D}}(r)$ implies that the set of discount factors that can support collusion is smaller. Now, we are ready to compare our two data sharing policies to determine which is more sustainable for collusive pricing.

Lemma 1 *Given $\delta \geq \hat{\delta}^{mkt}_{\mathcal{D}}$, the platform's optimal ad valorem fee to sustain collusion under marketplace mode is:*

$$r^{mkt}_{\mathcal{D}}(\delta) = \begin{cases} 1 - \frac{\delta u}{\delta(1-\frac{\tau}{2})-\frac{1}{2}+\frac{3\tau}{8}} & \text{if } \mathcal{D} = S \\ 1 - \frac{\delta u}{\delta(1-\frac{3\tau}{2})-\frac{1}{2}+\frac{5\tau}{4}} & \text{if } \mathcal{D} = N \end{cases}$$

where $\hat{\delta}^{mkt}_S = \frac{\frac{1}{2}-\frac{3\tau}{8}}{1-\frac{\tau}{2}-u}$ and $\hat{\delta}^{mkt}_N = \frac{\frac{1}{2}-\frac{5\tau}{4}}{1-\frac{3\tau}{2}-u}$. Otherwise, $r^{mkt}_{\mathcal{D}}(\delta) = 0$.

Lemma 1 characterizes the fee set by A to ensure collusion can be realistically sustained. The reason we preface that the δ be sufficiently large is that for low δ, collusion cannot be sustained. In such a case, we simply just set the optimal fee to 0. The specific thresholds for δ in Lemma 1 is derived by restricting r between 0 and 1. The properties of the fee are quite intuitive. Regardless of data sharing policy, the optimal fee is weakly increasing in δ and decreasing in u. If sellers are more patient or have lower outside options, the platform can extract more seller profit by setting a higher fee under collusion. One last immediate observation is that for less patient sellers, $r_N^{mkt} \geq r_S^{mkt}$ and vice versa.

Proposition 1 *Under marketplace mode, collusion is more sustainable with no data sharing than with full data sharing.*

When the platform operates under marketplace mode, sellers are more willing to collude with no data sharing. Because the platform always extracts profits from the seller until they are at u profit in competition, collusion sustainability is essentially dependent on the magnitude of deviating profits compared to collusive profits. Although the collusive profits increase with data sharing as sellers can set personalized prices for each consumer, the deviating profits increase even more as each seller can capture the entire market by setting price to every consumers' willingness-to-pay. Instead, if a seller deviates under no data sharing, they must also lower prices for consumers they would have won under collusion. Thus, deviating is less lucrative under uniform pricing. So far, our analysis has only considered how collusive outcomes given some exogenous data sharing policy. But how will endogenizing the benchmark policies affect A's decision?

Corollary 1 *Under our benchmark cases, the following summarizes A's choice of data sharing in marketplace mode:*

- *If $\delta < \underline{\delta}_N$, A does not share data and there will be no collusion;*
- *If $\underline{\delta}_N \leq \delta < \underline{\delta}_S$, A does not share data and there will be collusion;*
- *If $\underline{\delta}_S \leq \delta$, A shares data and there will be collusion*

where $\underline{\delta}_N$ is such that $r_N^{mkt}(\underline{\delta}_N) = \frac{\tau - 2u}{1 - \frac{\tau}{2}}$ and $\underline{\delta}_S$ is such that $\frac{r_S^{mkt}(\underline{\delta}_S)}{r_N^{mkt}(\underline{\delta}_S)} = \frac{4-2\tau}{4-\tau}$.

As mentioned earlier, the platform prefers not to share data in a competitive market. Accounting for this fact, our results from Proposition 1 still do not change. Indeed, collusion is more sustainable with no data sharing even when endogenizing our extreme data sharing policies. In a cartelized market, the platform will share data if sellers are patient enough. Recall from the discussion of Lemma 1 that the fee with data sharing exceeds that without data sharing. Coupled with the fact that given the same fee collusive profits are higher under no data sharing for sellers (i.e. $\pi_S^c(r) > \pi_N^c(r)$), as δ increases the platform will eventually prefer sharing data.

5 Equilibrium in the integrated case

In this section, we alter the market structure of the platform in the stage game. Platform A is now fully integrated with seller B_1, i.e. A now operates in dual mode. To simplify notation, we now denote seller B_2 as just B. Throughout this section, we refer to A as both a "platform" and "seller". The timing and all other assumptions of the game remain the same. Given that the sellers (A and B) are not symmetric as opposed to in Section 4, players will now follow different critical discount factors. We discuss how this slightly changes the analysis in Section 5.3.

5.1 Full Data Sharing, $\mathcal{D} = S$

If A shares all the data with B, both players can set personalized prices for all consumers. First, observe that since A and B compete for consumers through localized Bertrand competition, competing price offers are driven down to marginal cost. In equilibrium, A is willing to decrease prices up to its effective marginal cost $rp_B(x)$ and a consumer located at $x \in [0, 1]$ is indifferent between A and B iff

$$1 - rp_B(x) - \tau x = 1 - (1 - x)\tau - p_B(x).$$

Rearranging terms and accounting for the fact that $p_B(x) \leq 1 - (1 - x)\tau$, the optimal price B sets in equilibrium is

$$p^*_{B,S}(x, r) = \begin{cases} 0 & \text{if } x < \frac{1}{2} \\ \frac{(2x-1)\tau}{1-r} & \text{if } \frac{1}{2} \leq x < x^d(r) \\ 1 - (1 - x)\tau & \text{otherwise} \end{cases} \tag{8}$$

where

$$x^d(r) \equiv \min\{\frac{1 - r(1 - \tau)}{\tau(1 + r)}, 1\}. \tag{9}$$

Note that B sets price down to 0 for consumers who prefer A's product. This follows the conventional logic of Bertrand competition as marginal cost is 0. However, for consumers who prefer B's product, A and B will price compete until the price is lowered such that A is willing to forgo the demand. Indeed, at a sufficiently low price, A's gain from pricing lower is lower than the gain from just letting B win the consumer and obtaining a proportion of the profit. $x^d(r)$ represents the point in which B is starting to get price capped at $1 - (1 - x)\tau$. One important thing to note is that $x^d(r) \to \frac{1}{2}$ as $r \to 1$ and $x^d(r) \to 1$ as $r \to 0$. This is quite intuitive as if r goes to 1, A will never price compete for the consumer (that prefers

B's product) if $p_B(x) = 1 - (1-x)\tau$. Similarly, if r goes to 0, the environment is just like that of Thisse and Vives (1988). Without r, we can just interpret A as a competing seller, instead of a dual-role platform.

Now, consider the consumers who prefer A's good. Because B is willing to price at 0, A will set prices such that a consumer is indifferent between A and B. This occurs iff

$$1 - p_A(x) - \tau x = 1 - (1-x)\tau.$$

Thus, the optimal price A sets in equilibrium is

$$p^*_{A,S}(x,r) = \begin{cases} (1-2x)\tau & \text{if } x < \frac{1}{2} \\ r p^*_{B,S}(x,r) & \text{otherwise} \end{cases}. \tag{10}$$

This yields the following competitive payoffs for A and B:

$$\pi^*_{A,S}(r) = \int_0^{\frac{1}{2}} (1-2x)\tau dx + \frac{r}{1-r}\pi^*_{B,S}(r)$$
$$\pi^*_{B,S}(r) = \int_{\frac{1}{2}}^{x^d(r)} (2x-1)\tau dx + (1-r)\int_{x^d(r)}^{1} (1-(1-x)\tau)dx.$$

Lemma 2 *The three following properties hold: (i)* $\frac{d\pi^*_{A,S}(r)}{dr} > 0$*; (ii)* $\frac{d\pi^*_{B,S}(r)}{dr} \leq 0$*; (iii)* $\pi^*_{B,S}(r)$ *is weakly concave in* r.

Let R^*_S denote the largest ad valorem fee such that $\pi^*_{B,S}(R^*_S) = u$. Lemma 2 implies that R^*_S is the profit-maximizing fee for A as A's equilibrium profit is increasing in r and B's equilibrium profit is weakly decreasing in r.

Under full collusion, A and B maximize joint profits in the market. A and B achieve this by only attracting consumers who prefer their product by setting their price schedule to that of (2) and obtaining the collusive payoffs

$$\pi^c_{A,S}(r) = (1+r)(\frac{1}{2} - \frac{\tau}{8}), \tag{11}$$
$$\pi^c_{B,S}(r) = (1-r)(\frac{1}{2} - \frac{\tau}{8}). \tag{12}$$

Given the collusive price schedules, B's deviating strategy would be to undercut A's price for all consumers. A's deviating price would be the same, but only when the undercut price is greater than $rp_S^c(x)$. To summarize, A and B's deviating prices are

$$p_{A,S}^d(x,r) = \begin{cases} 1-\tau x & \text{if } x \leq x^d(r) \\ 1-\tau x^d(r) & \text{otherwise} \end{cases} \text{ and } p_{B,S}^d(x) = 1-\tau(1-x).$$

Deviating profits under full data sharing would then be

$$\pi_{A,S}^d(r) = \int_0^{x^d(r)} (1-\tau x)dx + r\int_{x^d(r)}^1 (1-\tau(1-x))dx, \tag{13}$$

$$\pi_{B,S}^d(r) = (1-r)\int_0^1 (1-\tau(1-x))dx = (1-r)(1-\frac{\tau}{2}). \tag{14}$$

Notice that collusive profits and deviating profits for B are exactly that from the non-integrated case in Section 4.1.

5.2 No Data Sharing, $\mathcal{D} = N$

If A is not allowed to share any data with B, seller B sets a uniform price for all consumers and only A can set personalized prices. To understand how we can arrive at a pure strategy equilibrium, consider our environment but without the two crucial assumptions: (i) the seller has an outside option of $u > 0$ and (ii) the platform prefers to keep the seller on the platform over selling on its own. This is akin to standard environments presented in the relevant literature. As previously mentioned, A and B compete for consumers through localized Bertrand competition. Thus, given some price p_B set by B, A's optimal strategy is to set $p_A(x) \geq rp_B$ such that consumers are indifferent between $p_A(x)$ and p_B. Specifically, A's optimal price schedule would be

$$\tilde{p}_A^*(x,p_B,r) = \begin{cases} (1-2x)\tau + p_B & \text{if } x \leq \tilde{x}(r,p_B) \equiv \min\{\frac{1}{2} + \frac{(1-r)p_B}{2\tau}, 1\} \\ 1-\tau x & \text{otherwise} \end{cases}. \tag{15}$$

However, B has an incentive to deviate to $p_B - \epsilon$ for some small ϵ and capture $(p_B - \epsilon)\tilde{x}$ in profit. This process will constantly continue until p_B is driven down to 0, but then B now has an incentive to increase prices above 0 as A is giving up half the market.

Lemma 3 *There exists a pure strategy equilibrium when $\mathcal{D} = N$. The equilibrium ad valorem fee and prices for A and B are characterized as follows:*

$$R_N^* = 1 - \frac{u}{1-\tau},\ p_{A,N}^*(x) = 1-\tau x,\ \ p_{B,N}^* = 1-\tau.$$

The logic of Lemma 3 goes as follows. When products are sufficiently similar, i.e. $\tau < \frac{1}{2}$, B is always willing to lower prices to attract the entire market demand. At a certain point, A is no longer willing to price compete and lets B have the market. However, A will recalibrate its ad valorem fee such that B's profit will be exactly u. In other words, B prices like a monopolist and A will extract profits up to the point that B still stays on the platform. With the added assumption, A is no longer willing to undercut B. Indeed, if A prefers to keep A on the platform and B has some positive outside option, there is an effective "stopping point" at which A gives up price competing. It is easy to see that A will set the highest r such that $\pi^*_{B,N}(r) = u$. More specifically, A will set $R^*_N = 1 - \frac{u}{1-\tau}$ which gives us $\pi^*_{A,N}(R^*_N) = 1 - \tau - u$ and $\pi^*_{B,N}(R^*_N) = u$. Lemma 3 also implies that the equilibrium profits for A and B, respectively, are

$$\pi^*_{A,N}(r) = r(1-\tau) \text{ and } \pi^*_{B,N}(r) = (1-r)(1-\tau).$$

Lemma 4 *Under no data sharing, collusive prices are as follows:*

$$p^c_{A,N}(x) = 1 - \tau x \textit{ and } p^c_{B,N} = 1 - \frac{\tau}{3}.$$

Consequently, the deviating prices are:

$$p^d_{A,N}(x,r) = \begin{cases} 1 + \frac{2\tau}{3} - 2\tau x & \textit{if } \frac{2}{3} < x \le x^d_N(r) \equiv \min\{\frac{1-r}{2\tau} + \frac{2+r}{6}, 1\} \\ 1 - \tau x & \textit{otherwise} \end{cases} \quad \textit{and } p^d_{B,N} = p^*_{B,N} = 1 - \tau$$

where A only has an incentive to deviate if and only if $r < \frac{3-2\tau}{3-\tau}$.

Under full collusion, A and B maximize industry profits just like before. Because B is restricted to setting a uniform price, B cannot fully extract consumer surplus on their end, so A is incentivized to attract more consumers under no data sharing. However, at some point, consumers are too far to reach for A, so they prefer to let those consumers purchase from B and extract profits through their ad valorem fee. Indeed, decreasing the amount of information B has access to means that the market is less efficient in terms of assessing consumers' willingness-to-pay. Lemma 4 implies that the collusive payoffs for A and B are:

$$\pi^c_{A,N}(r) = \frac{2}{3} - \frac{2\tau}{9} + r(\frac{1}{3} - \frac{\tau}{9}), \tag{16}$$

$$\pi^c_{B,N}(r) = (1-r)(\frac{1}{3} - \frac{\tau}{9}). \tag{17}$$

Naturally, B's deviating price is to just set its price as if it was a monopolist under full

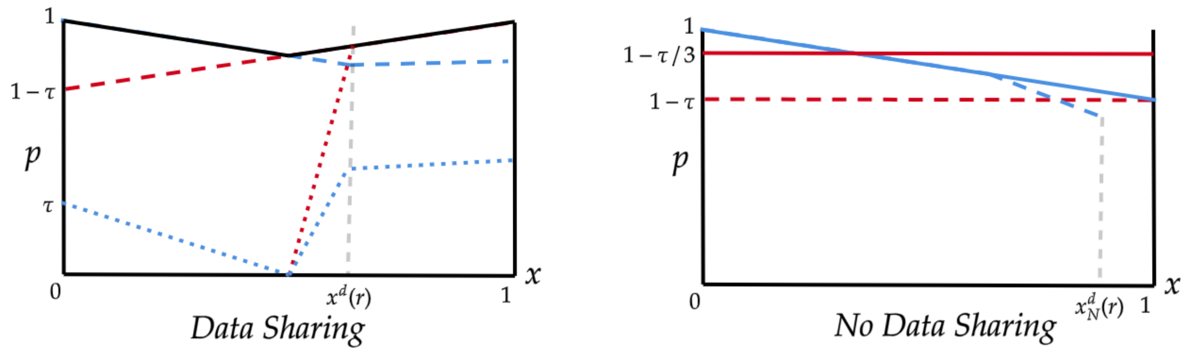

Figure 2: A graphical comparison of the price schedules based on different pricing policies under dual mode. Solid lines represent collusive prices, dashed lines represent deviating prices, and dotted lines represent competitive prices. Red lines represent the seller and blue lines represent the platform.

market coverage, i.e. $\tau < \frac{1}{2}$. A only has an incentive to deviate from the collusive outcome if the ad valorem fee is sufficiently low such that there is room to undercut B's uniform price. From Lemma 2, A and B's deviation payoffs are

$$\pi^d_{A,N}(r) = \begin{cases} \pi^c_{A,N}(r) & \text{if } r \geq \frac{3-2\tau}{3-\tau} \\ \frac{2}{3} - \frac{2\tau}{9} + \int_{\frac{2}{3}}^{x^d_N(r)} (1 + \frac{2\tau}{3} - 2\tau x)dx + r(1 - \frac{\tau}{3})(1 - x^d_N(r)) & \text{otherwise} \end{cases}, \tag{18}$$

$$\pi^d_{B,N}(r) = \pi^*_{B,N}(r) = (1-r)(1-\tau). \tag{19}$$

Notice that even though $\pi^d_{B,N}(r) = \pi^*_{B,N}(r)$, the profits will end up being different. As will be discussed later, the collusive ad valorem fee will be lower than that of the static Nash fee.

5.3 Comparison of Full Data Sharing vs. No Data Sharing

We say collusion is sustainable for player $j \in \{A, B\}$ under policy $\mathcal{D} \in \{S, N\}$ as an SPNE if and only if $\frac{\pi^c_{j,\mathcal{D}}(r)}{1-\delta} \geq \pi^d_{j,\mathcal{D}}(r) + \frac{\delta}{1-\delta}\pi^*_{j,\mathcal{D}}(r^*_{\mathcal{D}})$ where r is the ad valorem fee set under collusive pricing for policy $\mathcal{D}$. Let $\delta_{j,\mathcal{D}}(r)$ be the critical discount factor for player $j \in \{A, B\}$ under policy $\mathcal{D} \in \{S, N\}$ where

$$\delta^{dual}_{j,\mathcal{D}}(r) \equiv \frac{\pi^d_{j,\mathcal{D}}(r) - \pi^c_{j,\mathcal{D}}(r)}{\pi^d_{j,\mathcal{D}}(r) - \pi^*_{j,\mathcal{D}}(R^*_{\mathcal{D}})}. \tag{20}$$

Naturally, collusion is sustainable under policy k if it is sustainable for both players. In other words, δ must satisfy

$$\delta \geq \delta^{dual}_{\mathcal{D}}(r) \equiv \max_{j\in\{A,B\}} \delta^{dual}_{\mathcal{D}}(r) \text{ for } \mathcal{D} \in \{S, N\}. \tag{CR}$$

Like in the non-integrated case, $\delta_{\mathcal{D}}^{dual}(r)$ follows the same interpretation as a measure for collusive sustainability.

We first characterize the platform's maximization problem to ensure collusion is sustained given δ and $\mathcal{D} \in \{S, N\}$:

$$\max_{r} \pi_{A,\mathcal{D}}^{c}(r) \text{ s.t. } (*) \begin{cases} \pi_{A,\mathcal{D}}^{c}(r) \geq \pi_{A,\mathcal{D}}^{*}(r) & (IR_A) \\ \pi_{B,\mathcal{D}}^{c}(r) \geq u & (IR_B) \\ \delta \geq \delta_{A,\mathcal{D}}^{dual}(r) & (CR_A) \\ \delta \geq \delta_{B,\mathcal{D}}^{dual}(r) & (CR_B) \end{cases}.$$

The CR_A and CR_B constraints are merely derived from condition (CR). Because $\pi_{A,\mathcal{D}}^{c}(r)$ is increasing in r, we just need to find the highest r such that $(*)$ is satisfied. Solving for both policies gives us the following Lemma.

Lemma 5 *Given $\delta \geq \hat{\delta}_{\mathcal{D}}^{dual}$, the platform's optimal ad valorem fee to sustain collusion under dual mode is:*

$$r_{\mathcal{D}}^{dual}(\delta) = \begin{cases} 1 - \frac{\delta u}{(1-\frac{\tau}{2})\delta + \frac{3\tau}{8} - \frac{1}{2}} & \text{if } \mathcal{D} = S \\ 1 - \frac{\delta u}{(1-\tau)\delta + \frac{8\tau}{9} - \frac{2}{3}} & \text{if } \mathcal{D} = N \end{cases} \tag{21}$$

where $\hat{\delta}_{S}^{dual} = \frac{\frac{1}{2} - \frac{3\tau}{8}}{1-\frac{\tau}{2}-u}$ and $\hat{\delta}_{N}^{dual} = \frac{\frac{2}{3} - \frac{8\tau}{9}}{1-\tau-u}$. Otherwise, $r_{\mathcal{D}}^{dual}(\delta) = 0$.

Lemma 5 gives us A's optimal fee to ensure collusion can be sustained under dual mode. Observe that $r_{S}^{dual}(\delta) = r_{S}^{mkt}(\delta)$ as B's collusive, deviating, and competitive payoff functions are the exact same under marketplace mode. Under no data sharing, we instead find that $r_{N}^{dual}(\delta) \leq r_{N}^{mkt}(\delta)$. The increase in fee under dual mode comes from the information asymmetry between A and B. Under dual mode, A captures more of the market demand leaving B with less profit when colluding. Because B's competitive payoff remains at u regardless of mode and B's collusive payoff is greatly reduced (compared to its deviating payoff), B is less incentivized to collude now. A must then reduce its collusive ad valorem fee to ensure B cooperates.

Proposition 2 *Under dual mode, collusion is more sustainable with full data sharing than with no data sharing.*

The result in Proposition 2 is primarily driven by the seller B's incentives for collusion. Because B's off-path payoff of u remains the same regardless of data sharing, we can focus our attention to the collusive and deviating payoffs for both policies. As discussed in Lemma 5, the seller's critical discount factor given r is the exact same under data sharing regardless of mode. Now, recall that when the platform operated under marketplace mode with no data sharing, sellers had to internalize some profit loss from consumers on their turf to deviate from collusive pricing and capture the entire market. This same effect remains under dual mode, but is severely diminished as A now captures more of the market demand when fully colluding due to its informational advantage. Additionally, due to a decrease in collusive profit for the seller, the platform must now enforce a lower ad valorem fee to facilitate collusion. This reduction in fee also makes collusion less profitable for the platform.

Also, note that we know from the proof of Proposition 2 that A's fee in competition is lower under data sharing, i.e. $R_S^* < R_N^*$. This is due to A's willingness to give up the market under no data sharing and instead set a higher fee to extract seller rent. In fact, A's competitive profit is lower under data sharing. Therefore, sharing data also intensifies off-path competition which leads to a lower ad-valorem fee and profit under competition for A making collusion more lucrative.[9]

Corollary 2 *Suppose δ is sufficiently large such that collusion can be sustained for both players and A operates under dual mode. Then, collusion with full data sharing is more (less) profitable for A (B) than with no data sharing.*

Corollary 2 highlights an important point about sharing data under dual mode. The platform faces no draw backs as long as δ is sufficiently high. In contrast to marketplace mode where one policy might be more advantageous for collusion sustainability while the other may lead to higher collusive profits, collusion is both more sustainable and profitable under dual mode.

6 Extensions

6.1 Flexible Data Sharing

In this section, the platform can share data for any set of consumers with both sellers under marketplace mode. To be specific, A shares $(r, \mathcal{D}) \in [0,1] \times 2^{[0,1]}$ in period one of the stage game. In period two, sellers set personalized prices for $x \in \mathcal{D}$ and uniform prices for $x \notin \mathcal{D}$ where all pricing is still simultaneous. The rest of the stage game remains the same. Many of the insights introduced in the benchmark cases still hold, but this increase in flexibility makes the problem far more complex. As such, we will focus on how the platform can maximize collusive sustainability for individual sellers.

[9] Although for larger τ this does not hold, the aforementioned effect for the seller always holds and is the primary reason Proposition 2 holds.

Proposition 3 *In a competitive market, the platform will set $\mathcal{D} = \emptyset$ when operating under marketplace mode. In other words, the platform will not share any data with the sellers, even under flexible data sharing.*

Our result in Section 4 that competitive payoffs are higher for the platform without data sharing remains robust with the extra flexibility in sharing. Intuitively, sellers compete more fiercely for consumers they have data on. However, note that it may be possible that sharing certain subsets of consumers may inflate uniform prices from each seller. For example, if the platform sets $\mathcal{D} = [\frac{1}{4}, \frac{3}{4}]$, each seller will set their uniform price to $\frac{3\tau}{2}$ which is greater than when $\mathcal{D} = \emptyset$. We show that such an increase in profits from uniform pricing is overshadowed by the loss in profits from personalized pricing under any scenario where $\mathcal{D}$ is of positive measure.

Proposition 4 *In marketplace mode, collusion is most sustainable for sellers when $\mathcal{D} = \emptyset$.*

Similarly to before, restricting data from sellers maximizes collusive sustainability. Reducing sellers' critical discount factors involves finding what $\mathcal{D}$ makes deviating profits closest to collusive profits. Notice that in our benchmark cases, sellers' deviating prices were higher with data sharing implying that deviating gains are lower with no data sharing. This does not necessarily hold with more flexible sharing as the platform can share along the middle of the Hotelling line to reduce seller incentives to undercut with their uniform prices. However such a method to reduce deviating gains is met with even less deviating losses. Thus, the platform is just better off not sharing any data which would maximize deviating losses.

Proposition 5 *In dual mode, collusion is most sustainable for the non-integrated seller B when $\mathcal{D} = [\frac{1}{2}, \frac{2}{3}]$. In contrast, collusion is most sustainable for the platform A when $\mathcal{D} = \emptyset$.*

Under dual mode, the platform's optimal data sharing policy to facilitate collusion is less obvious. Without data sharing, the non-integrated seller only wins over consumers who strongly prefer their product under collusion, i.e. $x \in [\frac{2}{3}, 1]$. Therefore, to reduce deviating incentives it is best to "even the odds" by sharing data from $[\frac{1}{2}, \frac{2}{3}]$ to increase B's collusive yield. Just like in Proposition 4, if A wants to prevent B from setting a lower deviating uniform price, then A would need to share a sufficient amount of data along $[0, \frac{1}{2}]$. However, this would backfire and subsequently increase B's deviating price. The platform is then best off just sharing $\mathcal{D} = [\frac{1}{2}, \frac{2}{3}]$ to minimize B's critical discount factor.

As for the platform itself, not sharing any data is optimal for minimizing its own critical discount factor. Under such a scenario, the platform wins over $\frac{2}{3}$ of the consumers when colluding. Sharing any data in $[0, \frac{1}{2}]$ has no effect as the platform always has an information advantage and sets personalized prices for consumers on their half. In addition, sharing any data from $[\frac{1}{2}, \frac{2}{3}]$ only increases potential deviating gains. Finally, consider these last band of consumers from $[\frac{2}{3}, 1]$. Note that A only deviates until $x^d(r)$ with data sharing and until $x^d_N(r)$ without data sharing. We find that $x^d(r) > x^d_N(r)$ given that both these cutoffs are greater than $\frac{2}{3}$. Thus, not sharing data for these consumers is optimal for the platform. In conjunction, these conflicting results indicate that the optimal data sharing policy for collusive sustainability is a subset of $[\frac{1}{2}, \frac{2}{3}]$ depending on the values of our primitives. Indeed, restricting data may now be suboptimal for the platform to maximize collusive likelihood.

6.2 Privacy costs to consumers

Many recent data regulation efforts are primarily motivated by privacy concerns. For example, there has been a widely discussed debate on why consumers willingly provide websites with their personal information given their publicly expressed support for privacy (see Choi et al., 2019). To include a privacy concern for all consumers, we extend our model in the following way. If the platform shares data on the consumer $x \in [0, 1]$ with its sellers, then that consumer faces an additional "privacy" cost $c > 0$.

Suppose a platform operates in marketplace mode. This means that if the platform does not share data, consumers will have no concerns regarding privacy. The corresponding competitive, collusive, and deviating profits will follow exactly that of (5), (6), and (7). If the platform chooses $\mathcal{D} = S$, and assuming c is sufficiently small, the collusive and deviating prices for each seller will decrease by c. However, this change will not have enough bite to drastically affect the key insights highlighted in Section 4.3.

Instead, suppose a platform operates in dual mode. As the platform now acts as a seller and utilizes its data advantage, each consumer will face privacy cost c regardless of data sharing policy. We will abstract away from the competitive payoffs of the platform under $\mathcal{D}$ which we will denote as $U_{\mathcal{D}}$. This is because the competitive payoffs of the platform will not be very relevant in our proofs, but it is important to note that the payoffs will now be lower than before where there was no privacy cost, i.e., $U_S < \pi^*_{A,S}(R^*_S)$ and $U_N < \pi^*_{A,N}(R^*_N)$. With the added privacy cost, the non-integrated seller's collusive and deviating prices and the platform's collusive prices will lower by c. This means that under data sharing, both players will still split the market evenly, and without data sharing, the platform will still win over 2/3 of the consumers. However, the platform's deviating price not only decreases by c, but the set of consumers that the platform will capture when deviating also decreases. For example, when $\mathcal{D} = S$, both players will set their collusive price to $p^c_S(x) - c$ from (2). For

$x \in [\frac{1}{2}, 1]$, the platform will deviate iff

$$1 - \tau x - c \geq r(p_S^c(x) - c) \iff x \geq \frac{r\tau + (1-r)(1-c)}{\tau(1+r)}.$$

Thus, let

$$\tilde{x}^d(r,c) \equiv \min\left\{\frac{r\tau + (1-r)(1-c)}{\tau(1+r)}, 1\right\} \tag{22}$$

be the furthest consumer from A such that A will undercut B when deviating. It is immediate that $\tilde{x}^d(r,c) \leq x^d(r)$ from (9). We can similarly define

$$x_N^d(r,c) \equiv \min\left\{\frac{(1-r)(1-c)}{2\tau} + \frac{2+r}{6}, 1\right\} \tag{23}$$

as the furthest consumer that the platform captures when deviating. This implies that the platform's deviating profit given $\mathcal{D}$ is

$$\pi_{A,S}^d(r,c) = \int_0^{\tilde{x}^d(r,c)} (1 - \tau x - c)dx + r\int_{\tilde{x}^d(r,c)}^1 (1 - \tau(1-x) - c)dx, \tag{24}$$

$$\pi_{A,N}^d(r,c) = \begin{cases} \pi_{A,N}^c(r) - c\frac{2+r}{3} & \text{if } \tilde{x}_N^d(r,c) \geq \frac{2}{3} \\ \pi_{A,N}^c(r) - \frac{2c}{3} + \int_{\frac{2}{3}}^{\tilde{x}_N^d(r,c)} (1 + \frac{2\tau}{3} - 2\tau x - c)dx & \text{otherwise} \\ \qquad +r(1-\frac{\tau}{3})(1 - \tilde{x}_N^d(r,c) - c) & \end{cases}. \tag{25}$$

Proposition 6 *If consumers face a sufficiently low privacy cost $c > 0$, then the results in Propositions 1 and 2 still hold. That is, collusion is more (less) sustainable with no data sharing under marketplace (dual) mode.*

Proposition 6 provides a straightforward robustness result, but may have an important policy implication. Taking the contrapositive, if one were able to observe similar markets (aside from privacy costs) and find contradicting collusive outcomes, then privacy concerns from the platform may be a potential explanation. That is, consumers face varying privacy costs depending on separate markets which may be induced by factors such as platform data collection efforts and privacy transparency. Thus, if a given market is experiencing supra-competitive pricing and the other competitive pricing, then one plausible method to reduce collusive sustainability is to adjust the more collusive market's environment regarding personal information to be similar to the other.

6.3 Harshest Punishment

In this section, we briefly discuss why our previously established results remain robust if sellers use the harshest possible punishment strategy. The intuition is straightforward when the platform is in marketplace mode. Sellers can enforce the harshest punishment by setting their price to 0 and, regardless of data sharing policy, the platform will extract the other seller's profits to u. Therefore, the critical discount factors we characterize in Section 4.3 remain the same even with the harshest punishment strategies. Of course, we require that sellers must set nonnegative prices for this to hold. One immediate concern may be that our assumption that the platform is always motivated to ensure sellers have at least u profit is too restrictive. Even if we model the seller's entry decision on the platform, under the harshest punishment, both sellers would be left with u payoff, as the platform is incentivized to keep at least one seller (and fully extract down to u profit). If the other seller leaves, then their profit will be their outside option u, so the critical discount factors in Section 4.3 are unaffected.

Proposition 7 *If sellers enforce the harshest punishment, the result in Proposition 2 still holds. That is, collusion is less sustainable with no data sharing under dual mode.*

When the platform integrates with one of the sellers, the same intuition holds if the non-integrated seller deviates. Indeed, the platform already enacts the harshest punishment in competition given that the seller stays. Again, if we decided to include seller entry, then the platform could try to punish the seller by setting r high enough such that the seller does not join. But, this would still leave the seller with u payoff from its outside option. If instead the platform deviates, then the seller would price at 0 to minimize the platform's payoffs. Under any data sharing policy $\mathcal{D}$, the platform would be left with $\frac{\tau}{4}$ profits. This change in off-path payoffs does not drastically change the analysis and results and the same intuition discussed in Section 5.3 holds.

7 Concluding remarks

In this paper we analyzed the intersection of several emerging concerns for online platforms: increased sharing of private consumer information to sellers and the growing use of pricing algorithms which may lead to autonomous algorithmic collusion. We use a stylized model of seller competition in a repeated game to analyze this problem. We conclude that collusion is more likely and profitable for the platform when sharing data under dual mode. In contrast, collusion is less likely and platform profits are ambiguous when sellers have access to data in a standard marketplace.

Therefore, the key message that arises from our results is that a blanket ban on sharing data can potentially be ineffective for consumers. Instead, one must assess the relationship between data sharing policy and platform structure on collusive outcomes. Naturally, this type of policy remedy requires a case-by-case analysis as opposed to a simple ban on data sharing to be effective.

The environment we examined is special. That is, our attention is restricted to two sellers and a market where total sales remain constant. Of course, many of the results rely on our critical assumptions in the stage game. However, these assumptions are required to keep pricing simultaneous. Otherwise, we cannot properly differentiate whether results are caused by a change in data sharing policy or timing structure. We conjecture that this is the primary reason why research bridging personalized pricing and collusion is less developed. Thus, our paper faces the trade-off in being more stylized and providing a reliable take on data sharing on collusive outcomes.

Our model can still be extended in many ways. First, we do not simulate our environment using pricing algorithms. Using Q-learning algorithms similar to Calvano et al. (2020) and Johnson et al. (2023), among others, is counterintuitive by nature as Q-learning algorithms have no knowledge of the economic environment and adapt by experimentation. We leave the task of simulating personalized pricing through reinforcement learning for future research. Also, platforms share data with all sellers in our model. One could allow platforms to choose different data sharing decisions for each seller when operating in marketplace mode and analyze how it affects collusive outcomes. This comes with the caveat that there will not be a pure strategy equilibrium in the stage game, so additional assumptions would have to be imposed to guarantee a tractable and reliable model. Lastly, we assume that each set of consumers is unique in each time period. A possible extension could be to consider returning consumers in which sellers can learn about consumer locations along the Hotelling line.

References

Anderson, Simon, Alicia Baik, and Nathan Larson (2023) "Price Discrimination in the Information Age: Prices, Poaching, and Privacy with Personalized Targeted Discounts," *The Review of Economic Studies*, 90 (5), 2085–2115, 10.1093/restud/rdac073.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello (2020) "Artificial Intelligence, Algorithmic Pricing, and Collusion," *American Economic Review*, 110 (10), 3267–97, 10.1257/aer.20190623.

Chen, Le, Alan Mislove, and Christo Wilson (2016) "An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace," in *Proceedings of the 25th International Conference on World Wide Web*, 1339–1349, 10.1145/2872427.2883089.

Choe, Chongwoo, Jiajia Cong, and Chengsi Wang (2024) "Softening Competition Through Unilateral Sharing of Customer Data," *Management Science*, 70 (1), 526–543, 10.1287/mnsc.2023.4689.

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim (2019) "Privacy and personal data collection with information externalities," *Journal of Public Economics*, 173, 113–124, https://doi.org/10.1016/j.jpubeco.2019.02.001.

Competition and Markets Authority (2018) "Pricing Algorithms: Economic working paper on the use of algorithms to facilitate collusion and personalised pricing."

Döpper, Hendrik and Alexander Rasch (2024) "Combinable products, price discrimination, and collusion," *International Journal of Industrial Organization*, 94, 103071, https://doi.org/10.1016/j.ijindorg.2024.103071.

Friedman, James W. (1971) "A Non-cooperative Equilibrium for Supergames," *The Review of Economic Studies*, 38 (1), 1–12, 10.2307/2296617.

German Monopolies Commission (2018) "Algorithms and Collusion," *Biennial Report*, `https://www.monopolkommission.de/images/HG22/HGXXII_Summary.pdf`.

Gradwohl, Ronen and Moshe Tennenholtz (2023) "Selling Data to a Competitor," `https://arxiv.org/abs/2302.00285`.

Hanspach, Philip, Geza Sapi, and Marcel Wieting (2024) "Algorithms in the marketplace: An empirical analysis of automated pricing in e-commerce," *Information Economics and Policy*, 69, 101111, https://doi.org/10.1016/j.infoecopol.2024.101111.

Harrington Jr, Joseph E. (2025) "A Critique of Recent Remedies for Third-Party Pricing Algorithms and Why the Solution is not Restrictions on Data Sharing," *Available at SSRN*, https://dx.doi.org/10.2139/ssrn.5190097.

Helfrich, Magdalena and Fabian Herweg (2016) "Fighting collusion by permitting price discrimination," *Economics Letters*, 145, 148–151, https://doi.org/10.1016/j.econlet.2016.05.024.

Hotelling, Harold (1929) "Stability in Competition," *The Economic Journal*, 39 (153), 41–57, http://www.jstor.org/stable/2224214.

Hu, Yihong, Guo Li, Mengqi Liu, and Shengnan Qu (2023) "Information Sharing and Personalized Pricing in Online Platforms," *Production and Operations Management*, 10591478231225178, 10.1177/10591478231225178.

Häckner, Jonas (1996) "Optimal symmetric punishments in a Bertrand differentiated products duopoly," *International Journal of Industrial Organization*, 14 (5), 611–630, https://doi.org/10.1016/0167-7187(95)00509-9.

Johnson, Justin P., Andrew Rhodes, and Matthijs Wildenbeest (2023) "Platform Design When Sellers Use Pricing Algorithms," *Econometrica*, 91 (5), 1841–1879, https://doi.org/10.3982/ECTA19978.

Klein, Timo (2021) "Autonomous algorithmic collusion: Q-learning under sequential pricing," *The RAND Journal of Economics*, 52 (3), 538–558, https://doi.org/10.1111/1756-2171.12383.

Liu, Qihong and Konstantinos Serfes (2007) "Market segmentation and collusive behavior," *International Journal of Industrial Organization*, 25 (2), 355–378, https://doi.org/10.1016/j.ijindorg.2006.05.004.

Miklós-Thal, Jeanine and Catherine Tucker (2019) "Collusion by Algorithm: Does Better Demand Prediction Facilitate Coordination Between Sellers?" *Management Science*, 65 (4), 1552–1561, 10.1287/mnsc.2019.3287.

Montes, Rodrigo, Wilfried Sand-Zantman, and Tommaso Valletti (2019) "The Value of Personal Information in Online Markets with Endogenous Privacy," *Management Science*, 65 (3), 1342–1362, 10.1287/mnsc.2017.2989.

Navarra, Federico, Flavio Pino, and Luca Sandrini (2024) "Mandated data-sharing in hybrid marketplaces," ZEW Discussion Papers 24-051, Mannheim, https://hdl.handle.net/10419/300680.

OECD (2017) "Algorithms and Collusion: Competition Policy in the Digital Age," *OECD Roundtables on Competition Policy Papers* (206), https://doi.org/10.1787/258dcb14-en.

——— (2018) "Personalised pricing in the digital era," *OECD Roundtables on Competition Policy Papers* (222), https://doi.org/10.1787/db4d9c9c-en.

Peiseler, Florian, Alexander Rasch, and Shiva Shekhar (2022) "Imperfect information, algorithmic price discrimination, and collusion," *The Scandinavian Journal of Economics*, 124 (2), 516–549, https://doi.org/10.1111/sjoe.12465.

Rhodes, Andrew and Jidong Zhou (2024) "Personalized Pricing and Competition," *American Economic Review*, 114 (7), 2141–70, 10.1257/aer.20221524.

S.3686 - 118th Congress (2024) "Preventing Algorithmic Collusion Act of 2024," `https://www.congress.gov/bill/118th-congress/senate-bill/3686`.

Shaffer, Greg and Z. John Zhang (2002) "Competitive One-to-One Promotions," *Management Science*, 48 (9), 1143–1160, `http://www.jstor.org/stable/822606`.

Thisse, Jacques-Francois and Xavier Vives (1988) "On The Strategic Choice of Spatial Price Policy," *The American Economic Review*, 78 (1), 122–137, `http://www.jstor.org/stable/1814702`.

Wang, Qiaochu, Yan Huang, and Param Vir Singh (2022) "Algorithms, Artificial Intelligence and Simple Rule Based Pricing," *SSRN Electronic Journal*, 10.2139/ssrn.4144905.

A Proofs.

Proof of Lemma 1.

Let $\hat{r}_{\mathcal{D}}$ denote the highest r such that $\pi^c_{\mathcal{D}}(r) = u$. Using the analysis in Section 4.1 and Section 4.2, we know

$$\delta^{mkt}_S(r) = \frac{(1-r)(\frac{1}{2}-\frac{3\tau}{8})}{(1-r)(1-\frac{\tau}{2})-u} \text{ and } \delta^{mkt}_N(r) = \frac{(1-r)(\frac{1}{2}-\frac{5\tau}{4})}{(1-r)(1-\frac{3\tau}{2})-u}.$$

It can be easily verified that $\delta^{mkt}_{\mathcal{D}}(r)$ is strictly increasing in r for $r < \hat{r}_{\mathcal{D}}$ within our parameter restrictions. For collusion to be sustainable under $\mathcal{D} \in \{S, N\}$, it must be that $\delta \geq \delta^{mkt}_{\mathcal{D}}(r)$. As A's profit is strictly increasing in r, A's optimal ad valorem fee to sustain collusion is to set r such that $\delta = \delta^{mkt}_{\mathcal{D}}(r)$. If the derived fee is greater than 1 or less than 0 (i.e. collusion is not sustainable), we just restrict the fee to 0 which gives the presented result.

Proof of Proposition 1.

For collusion to be sustainable under N, it must be that $\delta \geq \delta^{mkt}_N(r)$ and $2\frac{r}{1-r}\pi^c_N(r) \geq \tau - 2u$. This occurs for

$$\frac{\tau - 2u}{1-\frac{\tau}{2}} \leq r \leq r^{mkt}_N(\delta).$$

Note that the $r^{mkt}_N(\delta)$ is increasing in δ under our parameter restrictions, so the lowest δ such that collusion is sustainable is when the $r^{mkt}_N(\delta) = \frac{\tau-2u}{1-\frac{\tau}{2}}$. Solving this gives the lowest δ such that collusion is sustainable under no data sharing as

$$\underline{\delta}_N = \frac{2\tau - 0.5 - u - \frac{15}{8}\tau^2 + \frac{5\tau u}{2}}{u(\frac{5\tau}{2}-1)-(1-\frac{3\tau}{2})^2} < \delta^{mkt}_S(0) \tag{A1}$$

where $\delta^{mkt}_S(0)$ is a lower bound on how sustainable collusion is under full data sharing.

Proof of Corollary 1.

First, observe that the competitive payoff of A is higher under no data sharing. Therefore, the first item follows the proof of Proposition 1. The collusive profit for A is higher under S than N iff

$$2r^{mkt}_S(\delta)(\frac{1}{2}-\frac{\tau}{8}) > 2r^{mkt}_N(\delta)(\frac{1}{2}-\frac{\tau}{4}) \iff \frac{r^{mkt}_S(\delta)}{r^{mkt}_N(\delta)} > \frac{4-2\tau}{4-\tau}. \tag{A2}$$

Note that for $\underline{\delta}_N < \delta \leq \frac{\frac{1}{2}-\frac{3\tau}{8}}{1-\frac{\tau}{2}-u} = \hat{\delta}^{mkt}_S$ from Lemma 1, $r^{mkt}_S(\delta) = 0$, so (A2) cannot be satisfied, where $\underline{\delta}_N$ is derived in the proof of Proposition 1. It can be easily verified that for $\delta > \frac{\frac{1}{2}-\frac{3\tau}{8}}{1-\frac{\tau}{2}-u}$,

$\frac{r_S^{mkt}(\delta)}{r_N^{mkt}(\delta)}$ is strictly increasing in δ and as $\delta \to 1$, $\frac{r_S^{mkt}(\delta)}{r_N^{mkt}(\delta)} > 1 > \frac{4-2\tau}{4-\tau}$. Thus, there must exist some $\underline{\delta}_S$ such that for all $\delta > \underline{\delta}_S$, (A2) is satisfied which gives us items 2 and 3.

Proof of Lemma 2.

Observe that if $r < 1-\tau$, $\pi^*_{B,S}(r) = \frac{\tau}{4}$. Otherwise,

$$\begin{aligned}
\pi^*_{B,S}(r) =& \int_{\frac{1}{2}}^{x^d(r)} (2x-1)\tau dx + (1-r)\int_{x^d(r)}^{1} (1-(1-x)\tau)dx \\
=&\tau[(x^d(r))^2 + \frac{1}{4} - x^d(r)] + (1-r)(1-\tau)(1-x^d(r)) + (1-r)\tau(\frac{1}{2} - \frac{(x^d(r))^2}{2}) \\
=&\tau[\frac{(x^d(r))^2}{2} - \frac{1}{4}] + (1-x^d(r))(1-r) + \frac{r\tau}{2}(1-x^d(r))^2 \\
=&\frac{(1-r+r\tau)^2}{2\tau(1+r)} + (1-r)\frac{\tau+r-1}{\tau(1+r)} + r\frac{\tau+2r-2-r\tau}{2(1+r)} - \frac{\tau}{4} < \frac{\tau}{4}.
\end{aligned}$$

It can be then shown that

$$\frac{d\pi^*_{B,S}(r)}{dr} = \begin{cases} 0 & \text{if } r < 1-\tau \\ \frac{3-4\tau-2r+\tau^2-r^2}{2\tau(1+r)^2} & \text{otherwise} \end{cases} \leq 0$$

which gives us item (ii). For $r < 1-\tau$, the proof of (i) is trivial. Solving for (i) when $r \geq 1-\tau$, we have

$$\begin{aligned}
\frac{d\pi^*_{A,S}(r)}{dr} = \frac{\pi^*_{B,S}(r)}{1-r} + \frac{r\pi^*_{B,S}(r)}{(1-r)^2} + \frac{r\pi^{*\prime}_{B,S}(r)}{1-r} > 0 \iff \frac{\pi^*_{B,S}(r)}{1-r} + r\pi^{*\prime}_{B,S}(r) > 0 \\
\iff (1-r)(\underbrace{4r-(1+r)(1+r^2)+2\tau(1-r)}_{\equiv z(r)}) + 2\tau^2 r > 0.
\end{aligned}$$

Note that a sufficient condition for the statement to hold is to show that $z(r) > 0$ which can be easily verified. When $r \geq 1-\tau$,

$$\frac{d^2\pi^*_{B,S}(r)}{dr^2} = \frac{-4+4\tau-\tau^2}{\tau(1+r)^3} < 0.$$

Therefore, $\pi^*_{B,S}(r)$ is weakly concave and (iii) holds.

Proof of Lemma 3.

Consider B's optimal price when selling by itself (as a monopolist). A consumer purchases from B iff

$$1-(1-x)\tau - p_B \geq 0 \iff x \geq 1 - \frac{1-p_B}{\tau}.$$

Thus, B's profit maximization problem is $\max_{p_B}\{p_B \min\{1, \frac{1-p_B}{\tau}\}\}$. This is a standard monopoly problem which yields the optimal price $p_B^* = 1-\tau$.

Suppose A sets r. We first derive A's optimal price schedule $\tilde{p}_A^*(x, p_B)$ given a price p_B. Given p_B, a consumer is indifferent between purchasing from A and B iff

$$1 - p_A(x) - \tau x = 1 - p_B - \tau(1-x) \iff p_A(x) = (1-2x)\tau + p_B.$$

A will only set such a price if $p_A(x) \geq r p_B \iff x \leq \tilde{x}(r, p_B) \equiv \min\{\frac{1}{2} + \frac{(1-r)p_B}{2\tau}, 1\}$. Otherwise, A is willing to give up the consumer to B. This gives us $\tilde{p}_A^*(x, p_B, r)$ from (15). Also, observe that A's deviation must still guarantee that B's profit is at least u. That is, it must be that

$$\pi_{B,N} = (1-r)p_B(1-\tilde{x}'(r, p_B)) \geq u \iff \tilde{x}'(r, p_B) \leq 1 - \frac{u}{(1-r)p_B}$$

where $\tilde{x}'(r, p_B)$ represents the marginal consumer after A deviates. If $\tilde{x}(r, p_B) \leq \tilde{x}'(r, p_B)$, then $\tilde{p}_A^*(x, p_B^*)$ is A's optimal deviating price schedule. If $\tilde{x}(r, p_B) > \tilde{x}'$, then $\tilde{p}_A'^*(x, p_B^*)$ is A's optimal deviating price schedule where

$$\tilde{p}_A'^*(x, p_B^*) = \begin{cases} (1-2x)\tau + p_B & \text{if } x \leq \tilde{x}' \\ 1 - \tau x & \text{otherwise} \end{cases}.$$

Now, given A's optimal deviating price schedule, B has a profitable deviation from p_B^* to $p_B^* - \epsilon$ where ϵ is near 0 to earn $(1-r)(p_B^* - \epsilon)$ profit. We can continue this process of price adjustments until B adjusts its price to $\underline{p}_B$ such that $(1-r)\underline{p}_B = u > 0$. A has no incentive to deviate and readjust its price schedule. By backwards induction, A sets the highest r such that $\underline{p}_B = \frac{u}{1-r} \leq 1-\tau$. That is, the optimal fee $R_N^* = 1 - \frac{u}{1-\tau}$ which implies $\underline{p}_B = p_B^* = 1-\tau$.

Proof of Lemma 4.

It is immediate that $p_{A,N}^c(x) = 1 - \tau x$ as consumer surplus is fully extracted for $x \leq \frac{1}{2}$. B sets a uniform price $p_B = 1 - \tau(1-\bar{x})$ where $\bar{x}$ is the marginal consumer that is indifferent between $p_{A,N}^c(x)$ and p_B. The problem boils down to:

$$\max_{\bar{x}} \int_0^{\bar{x}} (1-\tau x)dx + (1-\tau(1-\bar{x}))(1-\bar{x})dx.$$

Solving gives that the optimal $\bar{x} = \frac{2}{3}$ which gives us $p_{B,N}^c = 1 - \frac{\tau}{3}$. Given that $p_{A,N}^c(x) = 1 - \tau x$, it is obvious that $p_{B,N}^d = 1-\tau$, as when $\tau < \frac{1}{2}$, $1-\tau$ is the price B sets as a monopolist (see proof of Lemma 3). For A's deviating price, notice that A will keep its price of $1-\tau x$ for $x < \frac{2}{3}$. A can undercut consumers $x \geq \frac{2}{3}$ and set the highest price such that they are indifferent between buying from A and B which is satisfied iff

$$1 - p_A(x) - \tau x = 1 - (1-x)\tau - p^c_{B,N} \iff p_A(x) = 1 + \frac{2\tau}{3} - 2\tau x.$$

A is only incentivized to undercut B if

$$1 + \frac{2\tau}{3} - 2\tau x \geq r p^c_{B,N} \iff x \leq \frac{1-r}{2\tau} + \frac{2+r}{6}.$$

Thus, A is only incentivized to deviate if $\frac{1-r}{2\tau} + \frac{2+r}{6} > \frac{2}{3} \iff r < \frac{3-2\tau}{3-\tau}$.

Proof of Lemma 5.

Consider when $\mathcal{D} = N$. Fee r^{dual}_N satisfies the (IR_A) and (IR_B) constraints iff

$$1 - \frac{6\tau + 9u}{3-\tau} \leq r^{dual}_N \leq 1 - \frac{9u}{3-\tau}.$$

To satisfy (CR_B), it must be that

$$\delta \geq \delta^*_{B,N} = \frac{\pi^d_{B,N}(r^{dual}_N) - \pi^c_{B,N}(r^{dual}_N)}{\pi^d_{B,N}(r^{dual}_N) - \pi^*_{B,N}(R^*_N)} = \frac{(1-r^c_N)(\frac{2}{3} - \frac{8\tau}{9})}{(1-r^c_N)(1-\tau) - u} \iff r^{dual}_N \leq 1 - \frac{\delta u}{(1-\tau)\delta - \frac{2}{3} + \frac{8\tau}{9}}.$$

First, because $\delta^{dual}_{A,N}$ is weakly decreasing in r (see proof of Proposition 2), A will set $r^{dual}_N = \min\{1 - \frac{9u}{3-\tau}, 1 - \frac{\delta u}{(1-\tau)\delta - \frac{2}{3} + \frac{8\tau}{9}}\}$. Next, it is easy to verify that $r^{dual}_N = 1 - \frac{\delta u}{(1-\tau)\delta - \frac{2}{3} + \frac{8\tau}{9}}$ when $\delta \geq \frac{6-8\tau}{9-9\tau}$. When $\delta < \frac{6-8\tau}{9-9\tau}$, (CR_B) cannot be satisfied, so it is without loss to keep r^{dual}_N as $1 - \frac{\delta u}{(1-\tau)\delta - \frac{2}{3} + \frac{8\tau}{9}}$ as long as it is in [0,1]. For $\mathcal{D} = S$, we just use the proof of Lemma 1 as collusive, deviating, and competitive payoffs are the same for B as before in marketplace mode.

Proof of Proposition 2.

It is easy to verify that $\delta^{dual}_{A,N}(r)$ is continuous and differentiable for $r < \frac{3-2\tau}{3-\tau}$ and equal to 0 otherwise. For all other $j \in \{A, B\}$ and $\mathcal{D} \in \{S, N\}$, $\delta^{dual}_{j,\mathcal{D}}(r)$ is continuous and differentiable for $r \in (0,1)$. When $\delta^{dual}_{j,\mathcal{D}}$(r) is continuous and differentiable, observe that (20) implies

$$\begin{aligned} \frac{d\delta^{dual}_{j,k}(r)}{dr} &= \frac{\pi^{d\prime}_{j,k}(r) - \pi^{c\prime}_{j,k}(r)}{\pi^d_{j,k}(r) - \pi^*_{j,k}(R^*_k)} - \frac{\pi^d_{j,k}(r) - \pi^c_{j,k}(r)}{(\pi^d_{j,k}(r) - \pi^*_{j,k}(R^*_k))^2}\pi^{d\prime}_{j,k}(r) \leq (\geq)\, 0 \\ &\iff (1 - \delta^{dual}_{j,k}(r))\pi^{d\prime}_{j,k}(r) - \pi^{c\prime}_{j,k}(r) \leq (\geq)\, 0. \end{aligned}$$

It is easy to verify that $\frac{d\delta^{dual}_{A,k}(r)}{dr} \leq 0$ and $\frac{\delta^{dual}_{B,k}(r)}{dr} \geq 0$ given our parameter restrictions. When $r > \frac{3-2\tau}{3-\tau}$, $\delta^*_{A,N}(r) = 0$. Thus, $\delta^{dual}_{A,N}(r)$ is weakly decreasing. Moreover, it can be easily shown that $\delta^{dual}_{A,S}(r)$ is strictly decreasing and $\delta^{dual}_{B,S}(r)$ is strictly increasing for $r \in (0,1)$.

Lemma A.1 *The optimal ad valorem fee in the static Nash equilibrium under full data sharing* $R_S^* < 1-2u$ *which implies*

$$\delta_{A,S}^{dual}(r) < \tilde{\delta}_{A,S}^{dual}(r) \equiv \frac{\pi_{A,S}^d(r) - \pi_{A,S}^c(r)}{\pi_{A,S}^d(r) - \frac{\tau}{4} - (\frac{1}{2} - u)}.$$

Similarly, $\delta_{A,S}^{dual\prime}(r) < 0$.

Proof of Lemma A.1.

Evaluating $\pi_{B,S}^*(r)$ and $\frac{d\pi_{B,S}^*(r)}{dr}$ at $r = 1-2u$, we get

$$\begin{aligned}
\pi_{B,S}^*(1-2u) =& \frac{(2u+(1-2u)\tau)^2}{4\tau(1-u)} + \frac{u(\tau-2u)}{\tau(1-u)} + \tau(1-2u)\frac{(u\tau-2u)}{2\tau(1-u)} - \frac{\tau}{4} \\
=& \frac{4u\tau - 4u^2 + \tau^2 - 2u\tau^2}{4\tau(1-u)} - \frac{\tau}{4} < u \\
\iff& 4u\tau - 4u - \tau^2 < 0
\end{aligned}$$

which always holds. Because $\pi_{B,S}^*(r)$ is weakly decreasing in r and weakly concave, A must set $R_S^* < 1-2u$ to guarantee $\pi_{B,S}^*(r) \geq u$. The rest of the results in the lemma are immediate. ■

Lemma A.2 *The following two statements hold: (i) at* $\tilde{r}$ *such that* $\delta_{B,S}^{dual}(\tilde{r}) = \delta_{B,N}^{dual}(0)$, $\delta_{A,S}^{dual}(\tilde{r}) < \delta_{B,N}^{dual}(0)$ *and (ii)* $\min\{\delta_{A,S}^{dual}(0), 1\} > \delta_{B,S}^{dual}(0)$.

Proof of Lemma A.2. For the first item, $\delta_{B,S}^{dual}(\tilde{r}) = \delta_{B,N}^{dual}(0) = \frac{6-8\tau}{9(1-\tau-u)}$ when $\tilde{r} = r_S^c\big|_{\delta=\delta_{B,N}^{dual}(0)}$. That is,

$$\tilde{r} = 1 - \frac{\delta u}{(1-\frac{\tau}{2})\delta + \frac{3\tau}{8} - \frac{1}{2}}\Bigg|_{\delta=\frac{6-8\tau}{9(1-\tau-u)}} = 1 - \frac{(6-8\tau)u}{\frac{3}{2} - \frac{25}{8}\tau + \frac{5}{8}\tau^2 + (\frac{9}{2} - \frac{27}{8}\tau)u} > 1-\tau.$$

When $r = 1-\tau$,

$$\begin{aligned}
&\tilde{\delta}_{A,S}^{dual}(r) = 1 - \frac{\frac{1}{2} - \tau + \frac{\tau^2}{8} + u}{\frac{1}{2} - \frac{3\tau}{4} + u} < \delta_{B,N}^{dual}(0) = \frac{6-8\tau}{9(1-\tau-u)} = 1 - \frac{3-\tau-9u}{9-9\tau-9u} \\
\iff& (3-\tau-9u)(\frac{1}{2} + u - \frac{3\tau}{4}) < (9-9\tau-9u)(\frac{1}{2} + u - \tau + \frac{\tau^2}{8}) \\
\iff& u > \underbrace{\frac{24 - 86\tau + 75\tau^2 - 9\tau^3}{-48 + 46\tau + 9\tau^2}}_{<0 \text{ for } 0<\tau<\frac{1}{2}}
\end{aligned}$$

where $\tilde{\delta}_{A,S}^{dual}(r)$ is defined in Lemma A.1. By Lemma A.1, we have

$$\delta_{A,S}^{dual}(\tilde{r}) < \tilde{\delta}_{A,S}^{dual}(\tilde{r}) \leq \tilde{\delta}_{A,S}^{dual}(1-\tau) < \delta_{B,N}^{dual}(0) \;\Rightarrow\; \delta_{A,S}^{dual}(\tilde{r}) < \delta_{B,N}^{dual}(0).$$

For the second statement,

$$\delta_{A,S}^{dual}(0) > \frac{4-3\tau}{8(1-\frac{3\tau}{4})} > \delta_{B,S}^{dual}(0) = \frac{4-3\tau}{8(1-\frac{\tau}{2}-u)} \iff u < \frac{\tau}{4}. \blacksquare$$

As $\delta_{B,k}^{dual}(r)$ is weakly increasing in r, we know that the minimum δ such that collusion can be possibly sustained under no data sharing is $\delta_{B,N}^{dual}(0) = \frac{6-8\tau}{9(1-\tau-u)}$. Thus, by Lemma A.2, we know that as we go from $r = 0$ to the r such that $\delta_{B,S}^{dual}(r) = \delta_{B,N}^{dual}(0)$, the orders of the critical discount factor for A and B flip under data sharing. Also, because $\delta_{A,N}^{dual}$ ($\delta_{B,N}^{dual}$) is strictly decreasing (increasing) in r, the δ such that they intersect will be lower than $\delta_{B,N}^{dual}(0)$.

Proof of Corollary 2.

Recall that collusion can only be possibly sustained if $\delta > \frac{6-8\tau}{9(1-\tau-u)} > \frac{6-8\tau}{9-9\tau}$. Given our parameter restrictions, we first show

$$\frac{36-9\tau}{(1-\frac{\tau}{2})\delta + \frac{3\tau}{8} - \frac{1}{2}} < \frac{24-8\tau}{(1-\tau)\delta + \frac{8\tau}{9} - \frac{2}{3}} \tag{A3}$$

$$\iff 36(1-\tau)\delta + 32\tau - 24 - 9\tau\delta(1-\tau) - 8\tau^2 + 6\tau < 24\delta(1-\frac{\tau}{2}) + 9\tau - 12 - 8\delta\tau(1-\frac{\tau}{2}) - 3\tau^2 + 4\tau$$

$$\iff 12 + 25\delta\tau + 5\tau^2 - 5\delta\tau^2 - 12\delta - 25\tau > 0$$

$$\iff \underbrace{(1-\delta)}_{>0}\underbrace{(12-25\tau+5\tau^2)}_{>0} > 0.$$

Now, it is easy to directly show the two results.

$$\begin{aligned}
\pi_{A,S}^{c}(r_S^{dual}) > \pi_{A,N}^{N}(r_N^{dual}) &\iff r_S^{dual}(\frac{1}{2}-\frac{\tau}{8}) - r_N^{dual}(\frac{1}{3}-\frac{\tau}{9}) > \frac{1}{6} - \frac{7\tau}{72} \\
&\iff 12 - \tau + \delta u\left[\frac{24-8\tau}{(1-\tau)\delta + \frac{8\tau}{9} - \frac{2}{3}} - \frac{36-9\tau}{(1-\frac{\tau}{2})\delta + \frac{3\tau}{8} - \frac{1}{2}}\right] > 12 - 7\tau \\
&\iff 6\tau > \delta u\left[\underbrace{\frac{36-9\tau}{(1-\frac{\tau}{2})\delta + \frac{3\tau}{8} - \frac{1}{2}} - \frac{24-8\tau}{(1-\tau)\delta + \frac{8\tau}{9} - \frac{2}{3}}}_{<0 \text{ by (A3)}}\right]. \\
\pi_{B,S}^{c}(r_S^{dual}) < \pi_{B,N}^{N}(r_N^{dual}) &\iff (1-r_S^{dual})(\frac{1}{2}-\frac{\tau}{8}) < (1-r_N^{dual})(\frac{1}{3}-\frac{\tau}{9}) \\
&\iff 12 - \tau - \delta u\left[\frac{36-9\tau}{(1-\frac{\tau}{2})\delta + \frac{3\tau}{8} - \frac{1}{2}} - \frac{24-8\tau}{(1-\tau)\delta + \frac{8\tau}{9} - \frac{2}{3}}\right] > 12 - \tau \\
&\iff \frac{36-9\tau}{(1-\frac{\tau}{2})\delta + \frac{3\tau}{8} - \frac{1}{2}} - \frac{24-8\tau}{(1-\tau)\delta + \frac{8\tau}{9} - \frac{2}{3}} < 0.
\end{aligned}$$

Proof of Proposition 3.

Suppose A sets $\mathcal{D}$ such that joint seller profits are greater than under no data sharing. First, the personalized prices set by B_1 and B_2 follow exactly from the full data sharing case for $x \in \mathcal{D}$. Note that, as we know, profit under full data sharing is less than without data sharing, we assume A must share data for a positive measure of consumers to see if it is superior to no data sharing. Let $a \notin \mathcal{D}$ denote the farthest consumer from B_1 that purchases from B_1 and $b \notin \mathcal{D}$ denote the farthest consumer from B_2 that purchases from B_2 in equilibrium. It must be that $a \leq b$ given consumers only have unit demand. Therefore, we have two cases.

First, suppose $a = b$ which implies the consumer at location a is indifferent between purchasing from B_1 and B_2. This means that, in equilibrium, $1 - p_1 - \tau a = 1 - p_2 - \tau(1-a) \;\Rightarrow\; a = \frac{1}{2} - \frac{p_1 - p_2}{2\tau}$. B_1's maximization problem is

$$\max_{p_1} \left\{ p_1 \left[\frac{1}{2} - \frac{p_1 - p_2}{2\tau} - |\mathcal{D} \cap [0, a]| \right] \right\}.$$

Solving gives us $p_1^* = \frac{\tau + p_2}{2} - \tau|\mathcal{D} \cap [0, a]|$. Solving B_2's maximization problem similarly will give us $p_2^* = \frac{\tau + p_1}{2} - \tau|\mathcal{D} \cap [a, 1]|$. It is easy to verify that both p_1^* and p_2^* are less than τ if $\mathcal{D} \neq \emptyset$. Thus, it cannot be that $a = b$ and $\mathcal{D} \neq \emptyset$ such that sellers profit more than under no data sharing (implying the platform also cannot earn more profit).

Now, suppose $a < b$. Observe that this implies that $[a, b] \subseteq \mathcal{D}$. If there are consumers that are facing a uniform price in $[a, b]$ they are choosing not to purchase. So, the platform should share data for those consumers to gain additional profit. Denote $z_1 = |\mathcal{D} \cap [0, a]|$ and $z_2 = |\mathcal{D} \cap [b, 1]|$.

As such, the equilibrium uniform prices p_1^* and p_2^* must satisfy:

$$\frac{1}{2}(p_1^* - 2(b-a)\tau) = \tau(a - z_1) \quad \text{and} \quad \frac{1}{2}(p_2^* - 2(b-a)\tau) = \tau(1 - b - z_2).$$

The left hand side of each expression is the seller gains from deviating and undercutting from equilibrium pricing and the right hand side is the internalized loss from undercutting. Solving gives us $p_1^* = 2\tau(b - z_1)$ and $p_2^* = 2\tau(1 - a - z_2)$. Given z_1 and z_2, the platform should share $[0, z_1]$ and $[z_2, 1]$ as that gives the largest possible seller profit, so A's maximization problem boils down to:

$$\max_{a,b} a p_1^* + (1-b)p_2^* + \int_{\mathcal{D}} \max\{\tau(1-2x), \tau(2x-1)\} dx \;\Rightarrow\; a^* = b^* = \frac{1}{2}$$

which yields $z_1^* = z_2^* = 0$ and $a^* = b^* = \frac{1}{2}$ which is a contradiction. Thus, it must be that $\mathcal{D} = \emptyset$.

Proof of Proposition 4.

Observe that each seller's critical discount factor can be rewritten as

$$\delta_{\mathcal{D}}^*(r) = 1 - \frac{\pi_{\mathcal{D}}^c(r) - u}{\pi_{\mathcal{D}}^d(r) - u}$$

and is lower as $\pi_{\mathcal{D}}^c(r)$ is closer in value to $\pi_{\mathcal{D}}^d(r)$. Also, note that under full collusion, both sellers will always win on their respective halves.

Let our objective be to minimize B_1's discount factor to start. First, observe that it is optimal for A to share data in the form of $\mathcal{D} = [a, b]$ where $a \leq \frac{1}{2} \leq b$. This results in B_1 and B_2 setting collusive prices $p_1^c = 1 - \tau a$ and $p_2^c = 1 - \tau(1 - b)$, respectively.

To see this, suppose that A shared some data in the interval $[0, a]$, say $\mathcal{D}_1$ with positive measure z. To maximize collusion sustainability it is best to set $\mathcal{D}_1 = [a - z, a]$ as losses would be maximized because the new collusive uniform price would be greatest. Thus, our new shared data $\mathcal{D}' = \mathcal{D} \cup \mathcal{D}_1$ still holds the closed interval form. Now, suppose A instead removed some of the data they shared say $\hat{\mathcal{D}} \subseteq [a, \frac{1}{2}]$ from $\mathcal{D}$. If B_1's collusive uniform price remains unchanged, then potential gains from deviating may increase and if B_1's collusive uniform price decreases, then we can use the same principle mentioned before and set the new $\mathcal{D}'' = [0, a + |\hat{\mathcal{D}}|]$ which would maximize losses. Thus, the closed interval form will be preserved. Similar logic applies to seller B_2.

A's problem is to set a and b to minimize

$$\underbrace{\int_a^b (1 - \tau x)dx + \max\{(1 - \tau a)a, (1 - \tau(2 - b))(1 - b + a)\}}_{\pi_{\mathcal{D}}^d(r)} - \underbrace{\Big[\int_a^{\frac{1}{2}} (1 - \tau x)dx + (1 - \tau a)a\Big]}_{\pi_{\mathcal{D}}^c(r)}$$
$$= \int_{\frac{1}{2}}^b (1 - \tau x)dx + \max\{0, (1 - \tau(2 - b))(1 - b + a) - (1 - \tau a)a\}.$$

Solving gives us $a = b = \frac{1}{2}$ which means A shares data on measure zero of consumers. This is equivalent to setting $\mathcal{D} = \emptyset$.

Proof of Proposition 5.

Using similar logic from before, each player's critical discount factor is minimized when their collusive payoff is closest to their deviating payoff. First, consider minimizing B's critical discount factor. First, we show that B's deviating gains are minimized when $\mathcal{D} = [\frac{1}{2}, \frac{2}{3}]$. This is because under collusion with no data, B wins over consumers from $[\frac{2}{3}, 1]$, so sharing any data on consumers $x \in [\frac{1}{2}, \frac{2}{3}]$ will increase B's collusive payoff (hence decreasing deviating payoff). And if A shares any data on the interval $[0, \frac{1}{2}]$, B will set a higher deviating price $1 - \tau(1 - x)$ which is weakly greater than the uniform deviating price $1 - \tau$. Note that it is possible to share data in $[0, \frac{1}{2}]$ such that B does not deviate with its uniform price. However, this is only possible if A shares some $[a, \frac{1}{2}]$ such that

$$(1 - \tau)(\frac{1}{3} + a) \leq \frac{1}{3}(1 - \frac{\tau}{3}) \iff a \leq \frac{2\tau}{9(1 - \tau)}.$$

But, then B's deviating net gains is higher compared to just setting $\mathcal{D} = [\frac{1}{2}, \frac{2}{3}]$ as

$$\underbrace{\int_a^{\frac{1}{2}} (1-\tau(1-x))dx}_{\text{deviating net gains when } \mathcal{D}=[a,\frac{2}{3}]} \geq \frac{1}{2} - \frac{13\tau}{18} \geq \underbrace{\frac{2}{3}(1-\tau) - \frac{1}{3}(1-\frac{\tau}{3})}_{\text{deviating net gains when } \mathcal{D}=[\frac{1}{2},\frac{2}{3}]}.$$

Also, note that under this construction, losses are maximized as B can capture at most $\frac{1}{3}$ of the market with its uniform price under collusion.

Now, consider A's critical discount factor. If A shares data for consumers $x \in [\frac{1}{2}, \frac{2}{3}]$, then B will win the market under collusion. This means that A has possibility of deviating and undercutting these consumers, so to minimize deviating gains these consumers should not be included. It can be easily shown that when $x_N^d(r) > \frac{2}{3}$, $x^d(r) \geq x_N^d(r)$ and $p_{A,S}^d(x,r) \geq p_{A,N}^d(x,r)$. This means that when A is willing to deviate for a consumer $x \in [\frac{2}{3}, 1]$ under no data sharing, they are also willing to deviate under data sharing with a higher price. Thus, to minimize deviating gains none of the consumers' data from $[\frac{2}{3}, 1]$ should be shared.

Proof of Proposition 6.

Suppose $c < \tau$, the platform operatues in marketplace mode, and we are under full data sharing, i.e. $\mathcal{D} = S$. The collusive and deviating prices are just subtracted by c giving us the new collusive and deviating profits for the sellers:

$$\pi_S^c(r,c) = (1-r)(\frac{1}{2} - \frac{\tau}{8} - \frac{c}{2}) \text{ and } \pi_S^d(r,c) = (1-r)(1 - \frac{\tau}{2} - c).$$

The critical discount factor under full data sharing is then

$$\delta_S^{mkt}(r,c) = \frac{(1-r)(\frac{1}{2} - \frac{3\tau}{8} - \frac{c}{2})}{(1-r)(1-\frac{\tau}{2} - c) - u}.$$

Because this is increasing in r, the minimal bound for $\delta_S^{mkt}(r,c)$ is $\delta_S^{mkt}(0,c) = \frac{\frac{1}{2} - \frac{3\tau}{8} - \frac{c}{2}}{1 - \frac{\tau}{2} - c - u}$. Also, note that $\delta_S^{mkt}(0,c)$ is strictly decreasing in c as long as the denominator stays positive. Thus, using (A1),there must exist some bound $\bar{c}$ such that for $c < \bar{c}$, $\delta_S^{mkt}(0,c) > \underline{\delta}_N$.

Now, consider when A is in dual mode. Let $U_{\mathcal{D}}$ denote the competitive payoff of the platform in equilibrium. Note that $U_{\mathcal{D}} < \pi_{A,\mathcal{D}}^*(R_{\mathcal{D}}^*)$. Let $\delta_{j,\mathcal{D}}^{dual}(r,c)$ be the critical discount factor for player j given $\mathcal{D}$, r, and c. These critical discount factors are characterized as

$$\delta_{A,S}^{dual}(r,c) = \frac{\pi_{A,S}^d(r,c) - \pi_{A,S}^c(r) + \frac{1+r}{2}c}{\pi_{A,S}^d(r,c) - U_S}, \quad \delta_{A,N}^{dual}(r,c) = \frac{\pi_{A,N}^d(r,c) - \pi_{A,N}^c(r) + \frac{2+r}{3}c}{\pi_{A,N}^d(r,c) - U_N},$$

$$\delta_{B,S}^{dual}(r,c) = \frac{\pi_{B,S}^d(r) - \pi_{B,S}^c(r) - 0.5(1-r)c}{\pi_{B,S}^d(r) - (1-r)c - u}, \text{ and } \delta_{B,N}^{dual}(r,c) = \frac{\pi_{B,N}^d(r) - \pi_{B,N}^c(r) - \frac{2}{3}(1-r)c}{\pi_{B,N}^d(r) - (1-r)c - u}$$

using (11), (12), (13), (14), (16), (17), (18), (19), (24), (25). For future notational convenience, let

$$\tilde{\pi}^d_{A,S}(r,c) = \int_0^{\tilde{x}^d(r,c)} (1-\tau x)dx + r\int_{\tilde{x}^d(r,c)}^1 (1-\tau(1-x))dx$$

such that $\pi^d_{A,S}(r,c) = \tilde{\pi}^d_{A,S}(r,c) - c(1-\tilde{x}^d(r,c)) - rc\tilde{x}^d(r,c)$. The proof follows closely to Proposition 2. First, we can say that

$$\begin{aligned}
\delta^{dual}_{A,S}(r,c) &= \frac{\tilde{\pi}^d_{A,S}(r,c) - \pi^c_{A,S}(r) - c(1-r)(\tilde{x}^d(r,c)-\frac{1}{2})}{\tilde{\pi}^d_{A,S}(r,c) - rc\tilde{x}^d(r,c) - c(1-\tilde{x}^d(r,c)) - U_S} \\
&< \frac{\tilde{\pi}^d_{A,S}(r,c) - \pi^c_{A,S}(r)}{\tilde{\pi}^d_{A,S}(r,c) - c - U_S} \\
&< \tilde{\delta}^{dual}_{A,S}(r,c) \equiv \frac{\tilde{\pi}^d_{A,S}(r,c) - \pi^c_{A,S}(r)}{\tilde{\pi}^d_{A,S}(r,c) - c - \frac{\tau}{4} - \frac{1}{2} + u}.
\end{aligned}$$

The last inequality is implied by Lemma A.1 that $R^*_S < 1-2u$ and using the fact that U_S must be lower than the competitive payoff without privacy cost.

At the $\tilde{r}$ such that $\delta^{dual}_{B,S}(\tilde{r},c) = \delta^{dual}_{B,N}(0,c)$, $\tilde{r} < 1-\tau$ given sufficiently low c. This implies that

$$\begin{aligned}
\delta^{dual}_{A,S}(\tilde{r},c) < \tilde{\delta}^{dual}_{A,S}(\tilde{r},c) \leq \tilde{\delta}^{dual}_{A,S}(\frac{1-\tau-c}{1-c},c) \\
&= \frac{-\frac{\tau}{4} + \frac{\tau}{2(1-c)} - \frac{\tau^2}{8(1-c)}}{\frac{1}{2} - \frac{3\tau}{4} + u - c} \\
&< \delta^{dual}_{B,N}(0,c) = \frac{6-8\tau-6c}{9(1-\tau-u-c)}.
\end{aligned}$$

Also, $\delta^{dual}_{A,S}(0,c) > \frac{4-3\tau}{8(1-\frac{3\tau}{4})} > \delta^{dual}_{B,S}(0,c) = \frac{4-3\tau-4c}{8(1-\frac{\tau}{2}-u-c)}$ where the last inequality holds for sufficiently low $c > 0$. The proofs follow almost identically to the proofs for Lemmas A.1 and A.2. Thus, we can use the same argument in Proposition 2 yielding the same result.

Proof of Proposition 7

If sellers enforce the harshest punishment, then the non-integrated seller's critical discount factor do not change. However, if the platform deviates, then B will punish A by setting price to 0 and leaving A with $\frac{\tau}{4}$ profits regardless of $\mathcal{D}$. Thus, A's new critical discount factors are

$$\delta^{hp}_{A,\mathcal{D}}(r) = \frac{\pi^d_{A,\mathcal{D}}(r) - \pi^c_{A,\mathcal{D}}(r)}{\pi^d_{A,\mathcal{D}}(r) - \frac{\tau}{4}} < \delta^{dual}_{A,\mathcal{D}}(r).$$

Note that $\delta^{hp}_{A,\mathcal{D}}$ is still strictly decreasing in r. By Lemma **??**, we have

$$\delta^{hp}_{A,\mathcal{D}}(r) < \tilde{\delta}^{dual}_{A,S}(r) \equiv \frac{\pi^{d}_{A,S}(r) - \pi^{c}_{A,S}(r)}{\pi^{d}_{A,S}(r) - \frac{\tau}{4} - (\frac{1}{2} - u)}.$$

By the first statement of Lemma A.2, at $\tilde{r}$ such that $\delta^{dual}_{B,S}(\tilde{r}) = \delta^{dual}_{B,N}(0)$,

$$\delta^{hp}_{A,S}(\tilde{r}) < \delta^{dual}_{A,S}(\tilde{r}) < \delta^{dual}_{B,N}(0).$$

Also, we know that $\delta^{hp}_{A,S}(0) = \frac{4-3\tau}{8(1-\frac{3\tau}{4})} > \delta^{dual}_{B,S}(0)$ from the proof of the second statement in Lemma A.2. Thus, we can invoke the same argument in the proof of Proposition 2 to conclude that collusion is less sustainable with no data sharing under dual mode.

B Microfoundation for Assumption in the Stage Game.

We provide a brief stylized microfoundation for the assumption that the platform prefers to keep sellers rather than driving them out of the market. Suppose that there are two competing platforms, A_1 and A_2, that sellers can join and now there is a measure two of consumers. Before making their purchasing decision, consumers now must choose which platform to join. That is, we are assuming that sellers can multi-home and consumers are single-homing. Like the stage game, consumers have inelastic demand for one unit of the good and are differentiated in preference by seller. Now, consumers are also uniformly differentiated in preference for platform (i.e. $l_{A_1} = 0$ and $l_{B_2} = 2$). Specifically, consumer preferences are distributed independently and uniformly on $[0, 1] \times [0, 2]$. Crucially, consumers are "naive" in the sense that they make their homing decision without taking into account its effect on how sellers price on each platform. Thus, a consumer with platform preference $y \in [0, 2]$ is indifferent between joining A_1 and A_2 iff

$$-\tau_A y + \beta n_{A_1} = -\tau_A(2 - y) + \beta n_{A_2} \tag{B1}$$

where $\beta > 0$ represents the indirect network externality for consumers, τ_A is the transportation cost for choosing a platform, and n_i is the number of sellers on platform $i \in \{A_1, A_2\}$.[10] By (B1), the market size of platform A_1 and A_2 is $1 - \frac{\beta(n_{A_2} - n_{A_1})}{2\tau_A}$ and $1 + \frac{\beta(n_{A_2} - n_{A_1})}{2\tau_A}$, respectively. We keep all remaining assumptions on parameters which will be sufficient to guarantee players compete on the margin for their competitor's consumers. The timing of the stage game is now: (i) Platforms set their fees and data sharing policy; (ii) Sellers decide which platform(s) to join; (iii) Consumers decide which platform to join; (iv) Sellers set their respective prices; (v) Consumers observe prices and make their purchasing decision.

[10]We omit the value of the buyer's homing decision. As long as we keep the valuation uniform between both platforms, the condition will still be valid.

First, it is immediately obvious that under collusion, it is beneficial for the platform to keep both sellers. But will the platform prefer to host both sellers under competition? Suppose the platforms operate in marketplace mode and both platforms set their ad-valorem fee low enough such that both sellers join. Now, consider the case if A_1 deviates by setting a higher ad-valorem fee r_1 such that only one seller would join. Then the seller would earn

$$\Pi_{\mathcal{D}} = \begin{cases} (1-r_1)(1-\frac{\beta}{2\tau_A})(1-\tau_B/2) & \text{if } \mathcal{D}=S \\ (1-r_1)(1-\frac{\beta}{2\tau_A})(1-\tau_B) & \text{if } \mathcal{D}=N \end{cases}$$

where τ_B denotes the transportation cost for purchasing from sellers. It would then be optimal for A_1 to set the highest r_1 such that $\Pi_D = u$. That is,

$$r_1 = \begin{cases} 1-\frac{u}{(1-\frac{\beta}{2\tau_A})(1-\tau_B/2)} & \text{if } \mathcal{D}=S \\ 1-\frac{u}{(1-\frac{\beta}{2\tau_A})(1-\tau_B)} & \text{if } \mathcal{D}=N \end{cases}.$$

This leaves A_1 with the following competitive profit yield

$$\pi^*_{A_1} = \begin{cases} (1-\frac{\beta}{2\tau_A})(1-\tau_B/2)-u & \text{if } \mathcal{D}=S \\ (1-\frac{\beta}{2\tau_A})(1-\tau_B)-u & \text{if } \mathcal{D}=N \end{cases}.$$

Comparing this with how much the platform makes in equilibrium for Sections 4.1 and 4.2 gives us the following condition for this deviation to not be profitable:

$$\beta > \begin{cases} \frac{2\tau_A(1-\tau_B+u)}{1-\tau_B/2} & \text{if } \mathcal{D}=S \\ \frac{2\tau_A(1-2\tau_B+u)}{1-\tau_B} & \text{if } \mathcal{D}=N \end{cases}.$$

Therefore, if β is sufficiently high, or if the indirect network effect from sellers are strong enough, then the platform will opt to keep both sellers on the platform in equilibrium. Intuitively, this means that the platform values the seller's presence on the marketplace because consumers base their decisions on the number of sellers, e.g. product variety.